

22 July 1982

Who sells to the Soviet Union?

The Reagan Administration will probably decide to extend the grain agreement with the Soviet Union for a year. Can it then insist that West European governments should not buy gas from the East.

With the best will in the world, it will be hard for Europeans to be charitable about last week's special cabinet meeting called by President Reagan to help him decide whether United States farmers should continue to sell bulk cereals to the Soviet Union when present arrangements run out in September. Naturally, there will be much sympathy for Mr Reagan's dilemma. By denying it access to the Middle West, he could cause the Soviet Union a great deal of inconvenience (the need to shop around for other grain), physical discomfort (hunger) and even occasional starvation while establishing the powerful political point that United States farmers are more efficient than their Soviet counterparts. He would also provide the Soviet Union with a reminder that the *de facto* occupation of Afghanistan still rankles. These are great prizes, not lightly to be dismissed. But unfortunately for Mr Reagan, there is another millstone pressing on him — the unwillingness of farmers in the Middle West to sacrifice their occupations and their holdings by becoming bankrupt (more common already in the past twelve months than for several decades) in the cause of talking toughly to the Soviet Union. Politicians in most places will be full of fellow-feeling for the president. Especially because this is a mid-term election year (and because a decision must be made before polling day) they will weep with him as he sacrifices principles on the altar of prudence and renews the Soviet grain agreement for just one more year. But European politicians, no less given to calculations of how to square beliefs with the ballot box, will be asking why their acknowledgement of President Reagan's difficulty is not reciprocated in the matter of the natural gas pipeline between the Soviet Union and Western Europe.

That issue is more important than anybody in Washington seems to understand. (Mr George Shultz has made tactful noises, but is too new in his job as Secretary of State to be persuasive.) So far, three countries in Western Europe (France, Switzerland and West Germany) have committed themselves as customers for Soviet gas, and have agreed to make credits available for the construction cost of this gigantic project. (This in itself is not unusual; when did it most recently occur that Westinghouse, General Electric or Babcock-Wilcox built a nuclear power station in the hope of being able to sell the electricity production to some utility?) Sensibly enough, the countries concerned have done their best to ensure that they will not have to pay out too much (in the form of convertible currency) in advance by securing Soviet agreement to buy in much of the advanced equipment that will be needed. Washington has for the past six months been doing everything it can to ensure that such agreements will not be fulfilled — and that the Soviet Union will have to look to other suppliers or even make the equipment for itself. Since the summit meeting a few weeks ago at the palace of Versailles, the governments of Western Europe have been bubbling with resentment at President Reagan's Administration's ruling on this question. How can one government tell commercial corporations legally registered under another's laws that they should renege on what is otherwise a binding contract?

Western Europe's sin, on President Reagan's view, goes even deeper — its willingness to pay out convertible currency year after year in return for natural gas that may serve to keep its people warm and its industries in business but which will have the side-effect of helping the Soviet Union to pay its bills and will make

Western Europe strategically dependent on the East for a substantial part of its energy consumption. When Washington first woke up to the problems of the pipeline project eighteen months ago, it was principally alarmed at the prospect of strategic dependence, now mercifully seen in a less clear light. The present concern seems principally to deplore any arrangement that will make it less likely that the Soviet Union will cease to function overnight, and which may even offer it some assurance of an income-stream for thirty years or so. The long-term nature of the pipeline agreement is admittedly a more solemn undertaking than the one-year extension of the grain agreement that Washington now contemplates, but Western Europe could presumably repudiate its initial investment in the pipeline if the going became rough (or the price of gas too high). President Reagan is probably under-estimating the friendliness of his friends in Western Europe (one disservice) and is certainly bullying them too much by his complaints about the pipeline (another).

All this is especially disconcerting at a time when the Reagan Administration is in the box of having to talk to the Soviet Union, most days at Geneva, about the future of strategic arms. That contradiction is awkward and not easily avoided. Washington needs a Moscow willing to talk about strategic arms, but properly insists on its right to be huffy about Afghanistan and other transgressions. It wishes that it could also afford the luxury — all those farming votes — of standing firm on principle. Detente, the saying seems to be, is from time to time an expedient necessity. Western Europe, on the other hand, with no choice but that of being within a few minutes (by Intermediate Range Ballistic Missile) from the border, believes and behaves differently. Interdependence is a fact of life, but does not preclude the sudden breakdown of coexistence. The wisdom of the pipeline deal, in other words, is judged by relatively short-term criteria; for a half-continent whose separate existence could be quickly ended, long-term economic deals (lasting, say, thirty years) count for less than longer-term cultural connections — the sense of deprivation springing from people's inability to visit European cities such as Prague, Warsaw or even Leningrad when they choose.

Muddle about universities

The British government only seems to threaten university autonomy, but the precedent is bad.

Is the independence of the British university system, traditionally assured by the existence of the University Grants Committee as an intermediary between the government and the universities, now threatened? This is the obvious question provoked by last week's open letter to the committee from Sir Keith Joseph, the Secretary of State for Education and Science in the British government. For the letter, described as the first exchange in a "dialogue" between the government and the universities, finishes with the declaration that, in the development of British higher education and of the universities in particular, "there will be some strategic decisions for which it would be appropriate for ministers to take explicit responsibility and to answer in Parliament". After more than two years of abysmal relations between the government and the universities that it supports, beginning with the decision in the summer of 1979 that



thenceforth overseas students should pay arbitrarily defined "economic fees", and with the later and equally arbitrary decision that direct support for the universities must be reduced by 8.5 per cent, it is natural that the universities should assume that the minister's intentions towards the University Grants Committee are malevolent.

Some consideration should therefore be given to the more or less moderate tone of the bulk of the Joseph letter, which among other things repeats the promise that direct support for the universities will level out when the academic year 1983–84 is over (at least if inflation does not run away again), pats the grants committee on the back for its selective distribution of reduced funds among British universities, applauds the more recent decision to allocate extra funds to biotechnology (see *Nature* 20 May, p.173) and sympathizes with the general concern that, as things are, the recruitment of "young able people" into the university system is virtually impossible. The letter even acknowledges that the grants committee has acted sensibly in revising some of the targets for student numbers at particular universities in the light of some special pleading, even though these decisions will marginally shade the government's insistence that student numbers should be fixed from on high. How, it may be asked, can such a reasonable minister have malign intentions towards the independence of the system?

On the evidence of the minister's letter, part of the trouble is simply that the British government remains in a muddle about higher education. Student numbers are a good illustration. The most illiberal feature of the present university contraction is that each British university has been given a ceiling for the number of students from the United Kingdom and European Community countries it can have on its books in 1983–84. The result, since most numbers are reductions, is that much educational plant will be less than fully used and that universities will not be able to work their way out of financial trouble by doing their jobs more efficiently. (The threat that if they take in more than their allotted quota of students they will be penalized financially has not been withdrawn even though the grants committee did not repeat it in the letters sent to individual universities in May.) Yet this offensive system, and the uncertainties that go with it, arises simply from the government's fear that if the universities exceed their quotas, the cost will be increased of maintenance awards to students, paid by local authorities but then laid off on the central government. Is it really beyond the wit of a government anxious to play a more active part in the planning of higher education to devise some different system of subsidizing students, one that would combine the legitimate objective of controlling the total cost with flexibility on numbers?

The Joseph letter is also muddled about the relationship between the universities and the higher education institutions formally supported (but with funds from the central government) by local authorities, chiefly the polytechnics. The coexistence of these two kinds of institutions in what is called the "binary system" is strictly a political matter, for the polytechnics were established (in the 1960s) because of the then government's conviction that there was a need for a more down to earth system of higher education than the universities chose to provide. The experiment has been only a patchy success but, to its credit, the present government has been trying since before Sir Keith Joseph's time at the Department of Education and Science to wrest full control of the polytechnics away from the local authorities. Indeed, there is now a National Advisory Body for the polytechnics in England, and there may soon be one for those in Wales. The Joseph letter now asks that the grants committee and its counterpart committee for the polytechnics should consult with each other and that, as well as working out schemes for collaboration between individual universities and polytechnics, they should also help to define the relationship between the two sectors of higher education. The objective is sensible and should have been sought a long time ago. But the issue, eminently political as it is, must in the long run be settled by the government.

What the Joseph letter says about the support of university research is more alarming. The letter applauds the grants com-

mittee's intention to play its part in the dual support system, welcomes the Merrison report on the subject and goes on to approve of the principle that there should be "selectivity of funding" not merely by the research councils that provide project grants for university research but also by the grants committee, which is responsible for supporting a "basic level of research activity". What, suspicious minds will ask, is the minister approving? A decision, not yet taken, that in future there will be two kinds of universities — those equipped for research and those which are not? The Merrison committee rightly argued that British universities are less deliberate than they should be in using the grants committee's general subvention for research, and acknowledged that even if the job were done intelligently, some university departments would be able to mount research programmes only in collaboration with others. But the Merrison committee also rejected the principle of a close correlation between research support provided by the grants committee and the research councils, for that would only undermine the dual support system as a whole.

The fear that the Department of Education and Science does not properly understand the system of higher education in whose management it now seeks a more decisive role is the worry underlying the declaration that in future ministers must take more responsibility for it. Higher education may be too important to be left to academics, but the British government hardly seems their natural successor. By all accounts, the present intention is merely that Sir Keith Joseph should from time to time give the grants committee general guidance about its sense of where priorities lie. The difficulty is that the government's sense of priorities is unlikely to make academic sense. The present concern is somehow to shape higher education so as to contribute to national prosperity, a question with which university systems elsewhere than in Britain have wrestled for decades, but without much success. The danger is that from frustration, ministers will slip into the habit of making their priorities particular, issuing edicts such as "Let there be more engineers!", "Let there be more computer engineers!" or even "Let there be more computer engineers training on British equipment!" Sir Keith Joseph himself may be aware of the dangers. But once the precedent of intervention has been established, his successors will be provided with countless opportunities for foolish interference.

Pity the luckless British

*Great wars are said to bring out the best in people.
Small wars appear to reveal the worst.*

Who would have predicted that a mere month after a famous victory in the Falklands, the British government would be deeply embarrassed by a concatenation of domestic blunders — the false assumption that it would be able to comply with its bargain with its opponents when it won consent to send an amphibious force to the South Atlantic by mounting a Parliamentary enquiry on its own terms of reference? the discovery that an intruder could climb into Buckingham Palace and sit on the Queen's bed for ten minutes before being apprehended? and the growing likelihood that there has been another monumental breach of national security at one of the most technically advanced intelligence centres in the world? The British legend, especially common since the Second World War, is that the British are good at fighting (and even winning) wars but less endowed with flair in times of peace. The truth is different, as the events of the past few months have shown all too clearly. Most probably, the Parliamentary enquiry will show that the Falklands war was unnecessary, if not the consequence of a poor appraisal of intelligence then of indolence. Much the same is likely to be the case of the other conspicuous failures of public administration in the past few weeks. But who, or what, is to blame? The simple answer is the Civil Service, and its perennial pride in its technical innocence. The charge that this is so has been made repeatedly, and most governments have agreed that it is correct. When will one of them do something?

JPL takes on classified research

General to be appointed as new director

Washington

General Lew Allen Jr, who recently retired as chief of staff of the US Air Force, is reported to be the leading candidate for the directorship of the Jet Propulsion Laboratory (JPL), managed on behalf of the National Aeronautics and Space Administration (NASA) by the California Institute of Technology (Caltech). An announcement of the appointment is expected in the next few weeks.

The selection of General Allen would intensify the concern of some scientists at Caltech and JPL over the growing military role of the laboratory. Last year, the Caltech faculty and NASA agreed that up to 30 per cent of the laboratory's operating funds might come from Department of Defense (DoD) research contracts. Before that, the laboratory had refused to do classified research, under a policy laid down by Dr Bruce Murray (who ceased to be director of JPL on 1 July). The decline in NASA funding prompted Dr Murray to seek a change in that policy.

Dr Marvin Goldberger, president of Caltech, maintains that neither that change nor the appointment of General Allen — who he says is a "serious candidate" — will affect JPL's fundamental character. "We have every intention that JPL shall have its primary objectives in the civilian space programme. I can assure you that General Allen's view is coincident with this view."

A JPL scientist involved in the selection process pointed out that Allen has a strong technical and management background. He holds a PhD in physics from the University of Illinois, and worked at Los Alamos from 1954 to 1957. Allen directed the Air Force space programme from 1965 to 1972.

Allen declined to comment on the report that he is the principal candidate for the JPL position, and would say only that any announcement would come from JPL. But JPL officials are clearly worried about reaction to the appointment of General Allen, both within the Caltech-JPL community and outside it. Similar worries prompted the laboratory to adopt a tacit policy of not carrying out weapons research and development even with the lifting of the prohibition on classified work. The laboratory also made clear that it would not do "secret work secretly" — that is, it would only accept such classified projects as it could at least announce its participation in.

Dr Goldberger stresses, too, that work

for DoD is subject to scrutiny by a faculty committee which reviews its "propriety" and its effect on the "overall mission of the laboratory". He also insists "classified work will not jeopardize access to the laboratory by Caltech faculty and students".

Dr Arden Albee, JPL's chief scientist, agrees that these precautions are enough to

"I KNOW THE PAY'S GOOD—
IT'S THE NEW LAB-COATS
I DON'T LIKE."



ensure that JPL remains fundamentally a NASA laboratory. "Certainly it's something we have to worry about, but my own feeling is it's not going to cause a significant change in the character of the place", he said.

JPL's principal function within NASA is

France ponders biotechnology

France has done no microbiology for 20 years, has few people who understand fermentation or fungi and has never developed food processing technology. This wild exaggeration is the opinion of a francophile British biotechnologist sketching the problems facing France in his field. Now, however, a fairy godmother has waved her wand: the ministry of research and industry in Paris has published a vast programme for biotechnology, to stretch over three years and cost the government FF 600 million (about £50 million). Will she get Cinderella to the ball?

It depends in part on what the report says in a large section marked "Secret and Confidential", which has not been published but which describes the actions to be taken by the ministry in concert with industry. And it depends on whether France can create enough microbiologists, whose star fell faster and further in France than elsewhere when molecular biology turned to the eukaryotic genome. The University of Strasbourg, for example, was granted a new professorship in microbiology last year, with emphasis on biotechnology; it has been unable to fill it,

to provide the engineering and scientific support for unmanned planetary exploration missions. Some of this responsibility is shared with NASA's Ames and Langley laboratories, although JPL is the "lead centre".

At present, "about 8 or 9 per cent" of JPL's support comes from military contracts, according to Dr Albee. These are mostly "small technology projects" involving simulation, sensing and surveillance. The only large project is the autonomous spacecraft project which arose from the Voyager spacecraft work and is of some interest to the Air Force.

Nor does Dr Albee see any new threats to the scientific mission of JPL from its increased involvement in DoD projects: "NASA has always been a development agency; science got done on its periphery. That's as dangerous to science as DoD projects." He also points out that more research is supported by DoD on the Caltech campus than at JPL.

We have to put it in perspective," he said. "DoD is a major part of the budget. NASA is a small part."

JPL's current budget is \$321 million; the laboratory employs a full-time staff of 4,620. The proposed budget for 1983 is \$330 million, which represents a cut in real terms. According to Dr Arden Albee the most telling indicator of falling NASA support is that the laboratory has not started on a new planetary mission for almost 10 years. **Stephen Budiansky**

said a spokesman, with anyone with even a smattering of French.

The ministry's three-year programme promises to tackle the problem in partnership with the universities and the *grandes écoles*. And the Centre National de la Recherche Scientifique (CNRS), which supports much of the most advanced biology in France, plans to help double the number of French microbiologists.

Meanwhile, several important centres of biotechnology will now enjoy increased support. At Toulouse, for example, there are three laboratories grouped to form a kind of "transfer centre" between research and industry, directed by Professor Jean-Pierre Zaltar (a specialist in gene expression). At Marseilles, there is a laboratory for bacterial microbiology, which will be expanded next year with the help of Orsay biochemist Jean-Claude Patte. At Cadarache, there is a new laboratory for biomass studies run by CNRS in conjunction with the oil company Elf Aquitaine and the Commissariat à l'Energie Atomique. At Compiègne there is the University of Technology with Dr Daniel Thomas specializing in enzyme technology; in Paris, the Institut Pasteur is

working hard to make up for past years; and at the University of Strasbourg, there are well-advanced plans for a laboratory of plant genetics (which is, incidentally, strongly supported by the new report).

This week's report surprisingly emphasises the plant sciences, with no fewer than ten recommendations for INRA, the agricultural research body. Even Professor Roger Monnier, director of life sciences at CNRS, believes plant science must receive the most attention.

Over biotechnology as a whole, the report considers France to be weakest in bioengineering — in the kinetics of growth and production, in techniques of culture of microorganisms and cells, in enzymology, in reactor design, in extraction and purification and in the provision of analytical control equipment. France also suffers because of an "excessive compartmentalization and specialization of disciplines, research organizations and industries" says the report. But it ends on a strong note: now more than ever, France must support fundamental research, which is "the unique source of the unpredictable".

Robert Walgate

British Telecom Roll up!

The British government is embarking on the most significant phase in its plans to liberalize the telecommunications industry. Mr Patrick Jenkin, Secretary of State for Industry, said this week that he will be introducing legislation in November to sell off British Telecom (BT), the state-owned telecommunications monopoly. The subsequent sale of shares will be the "largest single issue on the London market and maybe the world".

The sale of British Telecom has been widely forecast. Earlier in the year it was suggested that the government was considering privatization to avoid acting as guarantor for the company's borrowings. This week, Mr Jenkin confirmed this. But the government clearly hopes that British Telecom will be able to keep down its customer charges, which have recently risen above the cost of services, simply to finance 90 per cent of the annual £2,200 million investment programme.

The proposal to sell British Telecom goes beyond measures earlier this year to liberalize British telecommunications. Last October, the government passed legislation allowing private companies to supply equipment for attachment to the network in competition with British Telecom. Equipment is still awaiting approval. Earlier this year, the Mercury Consortium, a group of three private companies, was licensed to operate a new telecommunications network in competition with British Telecom. And shortly, Mr Jenkin promises a general licence permitting the use of BT and Mercury networks by suppliers of value-added or

third party services.

The new bill will transform British Telecom into a public limited company and allow the sale of up to 51 per cent of the shares to the public. As soon as half the shares have been sold, the government will relinquish its control, allowing the company to borrow from shareholders and private markets. British Telecom's present status as a licensing authority will be transferred to the industry secretary acting through a newly-created Office of Telecommunications. Regulations — balancing the interests of those involved in the supply and use of telecommunication services — will be controlled by the new office. Licences to operate services will be issued, according to Mr Jenkin, only to those companies willing to fulfil their public duties by, for example, supplying uneconomic services to rural areas.

No doubt still smarting from the embarrassment of under-estimating the share price of Amersham International, sold off earlier this year, Mr Jenkin is as yet unwilling to hazard a guess at the value of British Telecom or the shares that will be on sale. Neither does he seem in any hurry to push the sale through. The legislation can be expected to be enacted by the end of next year, he says. But the sale of shares is not scheduled before the next election, which must occur before May 1984. The present government is clearly confident that it will have another term to run.

Judy Redfearn

Britain's nuclear power

PWRs hit snag

Mr Ron Anthony, recently appointed Chief Inspector of Nuclear Installations in Britain, wants to be counted on the side of the angels. Last week his Nuclear Installations Inspectorate (NII) published a report critical of several aspects of the "preconstruction safety report" for a British version of the Westinghouse pressurized water reactor (PWR), thus confounding environmentalists' fears that NII is in the pocket of the nuclear industry.

"As far as we're concerned, no-one could go too far in the matter of safety" said Mr Anthony last week, stressing that assessment of the Central Electricity Generating Board (CEGB)'s design would be continuous right up to the moment of operation. Next January, there will be a planning inquiry on the PWR, but even if the inquiry favoured the reactor, NII would still have the right to refuse a licence.

Is this unexpected opponent causing shudders in the nuclear industry? Not yet. Mr Anthony's inspectorial bark is judged to be worse than his bite. The director of CEGB, Sir Walter Marshall — nuclear physicist and passionate advocate of the PWR — said he was delighted that NII had flexed its muscles, while at the same time giving CEGB a pass mark on all the difficult issues (such as containment vessel

design). Moreover, the NII report says frequently that a satisfactory and safe PWR design is "achievable", and that CEGB has only to satisfy NII on certain points.

These are many, and include "human factors" now known to have played a great part in the Three Mile Island accident. "In the preconstruction safety report [for the British PWR] there is only brief mention of the role of the operator in testing and maintenance, operating procedures, operator training, the control room and operator response to faults" says the NII report. "The various sections give little indication of intent to design and operate on good ergonomic principles." NII is not unduly worried, however: CEGB is considering the problem, the inspectorate says, and NII will have evidence of the CEGB conclusion before a licence is granted.

More substantial, in NII eyes, are five concluding points gathered together as "not yet satisfactory". These are:

- Hazards presented by fire, earthquake and aircraft crash.
- "Ballooning" of the fuel cladding.
- Steam generator tube integrity.
- The automatic reactor protection system, which is a completely new design.
- CEGB's software models of fault development.

The report also says that the possible consequences of severe accidents will be the subject of a separate CEGB submission and NII report.

Of all these, Sir Walter Marshall is most worried about ballooning of the fuel cladding, a phenomenon discovered by UK Atomic Energy Authority researchers, Sir Walter's former colleagues. The problem is that uranium fuel rods necessarily contain a small amount of helium and that if the surrounding coolant pressure suddenly drops, this gas can balloon out of the cladding and close off coolant flow completely. For this to happen, the cladding ductility — and hence fuel-rod temperature — has to be just right, and the pressure must not fall too fast (otherwise the cladding simply bursts, allowing coolant past the ragged edges). The combination of conditions is unlikely, but it requires only a small loss-of-coolant accident, said Sir Walter. Experiments are under way in Canada to measure precisely the conditions in which ballooning occurs, but Sir Walter "stakes his reputation" in predicting that the phenomenon will prove to be no substantial obstacle to reactor safety.

The automatic reactor protection system is the other, more substantial, doubt. Relying on modern microelectronics and computer control, this system would handle major faults for 30 minutes before an operator was allowed to touch the controls. NII has no objection to the principle, but remains to be satisfied that it will work. That will require six years of software analysis and testing, said Sir Walter last week. He will set up a team of 10–20 specialists to do the job.

Robert Walgate

Whales' Waterloo

Science, it seems, has once again failed the sperm whale. Among other issues, the annual meeting of the International Whaling Commission (IWC), which convenes this week in Brighton, must decide on sperm whale catch limits. Quite how it will do so is not at all clear. The Scientific Committee of IWC, which met two weeks before the main commission, has, as last year, failed to make a recommendation.

Indecision continues to cast a cloud over the sperm whale. At a special meeting in March (*Nature* 1 April, p.383) the Scientific Committee, using the computer model developed by the International Institute of Environment and Development (IIED) and supported by all but the Japanese scientists, estimated the number of exploitable whales at around 200,000. The analysis indicated that even with a zero catch, stocks would continue to decline. The Japanese presented their own computer model and pushed for a minimum quota of 890. A decision on catch limits was deferred to this week's annual meeting.

The recent meeting of the Scientific Committee produced a few surprises. A majority of the committee agreed that the IIED model estimate of population size had been too optimistic and that the population continues to decline. The Japanese did not re-present their earlier model but put forward a modification of the IIED model indicating that earlier population estimates were low.

In view of the Japanese dissent, the Scientific Committee was unable to recommend a catch limit or to classify the stock. Because of the disagreement on the state of the sperm whale stock it is difficult to assess whether a catch of 890 could be sustained without taking a serious toll on the population. The buck has now been firmly passed back to the IWC delegates. A political trade-off must now be made. The sperm whale is not the species in greatest danger of extinction. With the conservationist countries determined to stand firm and press for a total ban, some clever manoeuvring is now called for if IWC is to avoid even greater schism.

Jane Wynn

The issue of gassing badgers with cyanide has been contentious since 1975, when the Ministry of Agriculture, Fisheries and Food began gassing badgers suspected of being carriers of bovine tuberculosis. The practice was suspended towards the end of 1979 after complaints from conservationists and wildlife organizations that it was unnecessary, inhumane and ineffectual, but resumed in October 1980 after a report from Lord Zuckerman that confirmed that badgers can be reservoirs for bovine tuberculosis, a conclusion disputed by some conservationists.

The Zuckerman report on badgers also pointed to the lack of experimental evidence of the effects of cyanide on badgers, when the study commissioned by the agriculture ministry from the Chemical Defence Establishment at Porton Down (not to be confused with the now-civil microbiology establishment at the same place). The report of that study, by anonymous officials at the establishment, was published last week.

The first objective of the study was to establish a kind of dose-response relationship for mortality among ferrets, chosen as

exposed to cyanide gas. The only badger that died was exposed to 6,560 mg min m⁻³ for 17 minutes; the other three have been returned to the wildlife park from which they came.

The government's decision to abandon the use of cyanide against badgers has been prompted by the knowledge that concentrations of cyanide in the tunnels of badger setts rarely approaches that needed to kill them within minutes, if at all. Its problem now is whether it can devise an alternative humane way of killing badgers — and do so quickly. Conservation interests will be able to exploit a prolonged hiatus in the control programme to reopen the question of its necessity. The use of carbon monoxide and nerve gases as alternative methods of control was ruled out in the Zuckerman report as being respectively ineffectual and "too dangerous".

British academic industry

Prize incentives

"We realize that we're just a football. We're here to be kicked by both sides." Thus a spokesman for the British Technology Group (BTG) on the group's present interest — the marriage of industry and academic science to put high technology on the market. In its latest venture, an academic enterprise competition, BTG hopes to encourage academics to don the entrepreneurial cap. Last week, Patrick Jenkin, Secretary of State for Industry, handed out cheques totalling £230,000 to twelve winning entries.

The academic enterprise competition, originally sponsored by the National Research Development Corporation (now part of BTG), was open to science researchers from British institutions of higher education who had set up, or planned to set up, a business based on their research work. At the award ceremony, enthusiasm ran high. Sir Alastair Pilkington, chairman of the panel of five judges, declared "We would like to see a large number of academics become millionaires every year". Patrick Jenkin described the occasion as one of "unique importance". The competition was all about the partnership between education and industry and aimed to dispel the feeling that "somehow for an academic to go into industry is disreputable". He issued a warning that the government's position was firmly on the sidelines. Public funds were but the seeds of support.

Professor W.A. Gambling of the University of Southampton, whose company York Technology Ltd took the top prize of £50,000, agreed that if Britain is to "seize the juicy plum" and commercialize high technology, higher education must play its part. But he added that the main occupation of the academic remains teaching and research.

Although legislation has yet to be passed, the National Research Develop-

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

experimental animals on the recommendation of Lord Zuckerman. One of the defects of the study is that the sample of ferrets (obtained from a commercial supplier) appeared to consist of two groups differing in body weight. The animals were exposed in groups of three to various concentrations for periods of 1, 5 and 25 minutes. Nothing is said in the report of the age of the animals, which had nevertheless been bred especially for the study.

The exposure of ferrets is measured as the product of the concentration of cyanide in the air and the time in minutes for which the ferrets breathed it. The ferret part of the study was used so as to estimate the likely lethal doses for badgers, fourteen times heavier than ferrets on the average, by means of the rule that the exposure required to produce a specified mortality is proportional to (body-weight)^{0.7}.

On this basis, the estimated exposures to cyanide needed to kill 90 per cent of badgers after times of 1, 5 and 25 minutes are 2,000, 5,000 and 5,500 mg min m⁻³ respectively. The report says that these "crude" estimates are supported by the experiments in which four badgers were

Eric Hoskins

Badger gassing

Ferrets stand in

The study which earlier in the month precipitated the British government's abandonment of the policy of gassing badgers appears to have provided the most detailed information so far of the mortality of ferrets exposed to hydrogen cyanide. A total of 177 female ferrets were exposed to hydrogen cyanide in the course of the study, but only four badgers were thus used because of their scarcity.

ment Corporation and the National Enterprise Board were merged to form BTG in September 1981. A Universities Coordination Unit was set up to go on the road and seek out entrepreneurs in the academic world. Last week's competition has at least provided a market survey. BTG now knows that there are at least 118 groups, the number of entries in the competition, keen to set up business ventures based on academic research.

Electronics and information technology dominated the entries but engineering, medical equipment and biotechnology were strongly represented. More than a third of the entries came from research institutions in the south of the United Kingdom. York Technology Ltd, the winner, was set up in 1980 to market products for the optical communications industry. Its first product, a "preform profiler" which measures the refractive-index distribution in glass preforms from which long lengths of optical fibre are drawn, was invented and patented at the University of Southampton. So far, £250,000 worth of equipment has been sold, 80 per cent as exports.

BTG may inject further cash, up to £250,000, in some of the prize-winning companies. It hopes, however, that private industry will become more interested and step in to bridge the investment gap. The competition seems to have established that while academics' sights may not yet be set on making a million, there is no shortage of people willing to give it a try. **Jane Wynn**

Synthetic fuels

Support wanes

Washington

The "synfuel fever" that led Congress in 1980 to launch a 12-year, \$88,000 million development plan seems to have lapsed into a synfuel coma. The US Synthetic Fuels Corporation, established by Congress to distribute the funds, is off to a slow start — several large privately-supported projects have recently collapsed and the economic projections of the profitability of synfuels are becoming bleaker.

Last week, the board of directors of the synfuels corporation continued at its measured pace, eliminating about half of the 35 proposals received in the second round of applications. The corporation has taken a very cautious course, so far selecting only two projects out of the first round and pledging to be "tough negotiators" in settling on the details of the government support for those. Chairman Edward Noble has also made clear that once the corporation has spent the \$14,600 million already appropriated to it, there will be no more funds available.

Yet even this is more than synfuel supporters had expected from the Reagan Administration. According to the Democratic staff of the House fossil and synthetic fuels subcommittee, the Administration now appears "to support at least the concept". But John Sawhill, who was Secretary of Energy and, briefly,

chairman of the synfuels corporation under President Carter, says that the corporation is doing "the minimum possible that will keep Congress off its back".

But critics of synfuel development say the industry's real problems go deeper than the policies of the synfuels corporation. Technological uncertainty, high risks and a declining world price of oil are formidable obstacles. Earlier this year, two large private projects collapsed, apparently for these reasons. Exxon, the world's largest corporation, pulled out of the Colony oil shale project in Colorado after the estimated construction costs doubled — to \$6,000 million.

The Alsands oil sands project in Alberta, Canada, suffered a similar fate, almost simultaneously, when both Shell Canada and Gulf Canada pulled out. And the two projects that the synfuels corporation has chosen to support — the Breckinridge oil-from-coal project in Kentucky and the Hampshire Energy gasoline-from-coal project in Wyoming — are reported to be having trouble finding equity backers.

An industry source said that Exxon's main trouble with the Colony project was that artificially high oil prices had made what are basically "neanderthal" technologies seem attractive.

Meanwhile, other projects are continuing. Union Oil is reported to be more than half-way to completion on its \$2,000 million, 10,000 barrel per day shale project in Colorado, and work is also progressing on the \$2,100 million Great Plains coal-gasification plant in North Dakota. Exxon, in spite of the setback on the Colony project, is going ahead with a small direct coal liquefaction plant in Texas and a coal gasification project in the Netherlands.

The industry is not dead; but neither is it just hibernating, as some of its supporters claim, waiting for the next rise in oil prices. The change of plan by Exxon was clearly a blow; as one observer put it, "it was a shock to everyone that Exxon couldn't do everything".

A reevaluation of the long-term economics of synfuels is at the heart of the industry's sickness. The draft of an Office of Technology Assessment study reports that even with the large loan guarantees available from the synfuels corporation, investment in particular projects seems attractive "only if current capital cost estimates for synfuels plants are correct and there are no cost overruns. Most industry experts, however, consider the chances of cost overruns to be high."

Some economists have also begun to question the rationale behind synfuel development. Arthur Wright of the University of Connecticut doubts that even a credible supply of synfuels at a credible price would insulate the American economy from the effects of oil price rises. "Synfuels might reduce imports, but not the source of the instability", he says.

Britain is odd-man-out on 2,4,5-T

The European Commission is urging the Community's member states to take greater precautions to ensure that herbicides containing 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) do not contaminate food. In a surprising move, the Commission is recommending that the Council of Ministers adopt a directive forbidding the use of 2,4,5-T on cereals and woodlands when fruits and mushrooms are in season and in weedkillers for domestic use. The Commission is also recommending that the maximum permitted level of the toxic contaminant 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) be reduced immediately to 0.01 mg per kg of 2,4,5-T and later to 0.005 mg per kg when a method of control at such a low level is available.

The Commission's move may embarrass the British government, even though Britain, together with Belgium and West Germany, already operates the 0.01 mg per kg limit for TCDD in 2,4,5-T. The British Ministry of Agriculture has repeatedly given 2,4,5-T a clean bill of health provided that it is used as recommended. The ministry's view was challenged in 1980 by the National Union of Agricultural and Allied Workers (NAAAW) which claimed to have

evidence of harm to individuals who had been in contact with 2,4,5-T. But the Advisory Committee on Pesticides, which the ministry asked to reply, saw little evidence to make it change its mind.

Thus, not surprisingly, NAAAW is greeting the Commission's move with enthusiasm while the agriculture ministry is more guarded, preferring to wait for the Council of Ministers' decision before responding to the Commission's recommendations.

The Commission's decision to move on the side of caution is the result of a review conducted in 1980 by its scientific committee for pesticides. Although that review found no evidence of direct harm from the proper use of herbicides containing 2,4,5-T, it nevertheless recommended lowering the level of TCDD in 2,4,5-T and greater precautions to minimize 2,4,5-T in food. Whether the Commission's recommendations become binding on EEC member states will depend on how they are greeted by the Council of Ministers. Whatever the outcome, however, Italy, the Netherlands, Denmark and West Germany will be unaffected. They have already banned use of the chemical as a herbicide.

Judy Redfearn

The industry's sorry state is compounded by Congress's loss of enthusiasm in the past two years. The 1980 Energy Security Act, which set up the government synfuels projects, also authorized spending on solar energy and conservation. With the dropping of all but the synfuels portion of the act, the coalition of environmentalists and oil-industry supporters that passed the act has fallen apart. A new anti-synfuels coalition, now made up of environmentalists and the much larger body of fiscal conservatives, is trying to destroy government synfuels support. So far, this coalition has had little success, but the days of government support for synfuels are surely numbered.

Stephen Budiansky

High-energy physics

Labour intensive

They have been playing musical chairs at CERN, the European Centre for Nuclear Research near Geneva. Last week there were six proposed experiments for LEP, the large electron-positron collider due to start up at the end of 1987. Now there are four, after a preliminary examination of the proposals gave a cautious thumbs down to the other two*.

But the number of physicists remains the same, so that the 850 scientists involved in all six proposals will attempt to regroup to find jobs for all, in a process that CERN dubs "optimalization". Final approval of the arrangements will be given at a research board meeting in November.

The scale of the LEP collaborations is thus going to be vast, far bigger than foreseen only two years ago. Then, the European Committee for Future Accelerators (ECFA), in a report on the future employment of physicists in Europe, predicted collaborations averaging only 75 physicists on each of six LEP experiments. Now the figures seem likely to reach 200 on each of four experiments. Rutherford, where is your string and sealing wax now?

Dr Jack Mulvey of the University of Oxford, who produced the ECFA report, said this week that he thought the main problem was going to be project management. But particle physicists had shown a remarkable ability to adapt to new conditions. "We've realized that it takes an enormous effort to get these big experiments on the floor", he said, and that large collaborations are necessary.

The LEP collaborations will also need to be big to spread costs — for this time CERN is not following its usual policy of offering about half the capital cost of the equipment, but will restrict itself to one-fifth. Since the total sum for the four experiments is some £75 million, the matter is not negligible.

The other problem is whether there will

be enough jobs for the 2,000 or so particle physicists in Europe when LEP is built. Mulvey thinks there will, as the ECFA report deemed the situation satisfactory with only 450 researchers at LEP. Although the intersecting storage rings at CERN will be closed down, and there will be less work at the Super-Proton Synchrotron, there will be enough to do at the proton-antiproton collider, at the low energy antiproton ring (LEAR) and at other European laboratories such as the Deutsches Elektronen-Synchrotron in Hamburg, the ECFA report concluded.

Although "four experiments" sounds rather modest, LEP should be an excellent physics factory, provided that the particle called the neutral intermediate vector boson (Z^0) exists. LEP is designed to be tunable to this resonance, which is the predicted mediating particle of the electro-weak interaction. If the resonance is not there, however, or if it proves difficult to tune to it, there will no doubt be 850 unhappy physicists.

Robert Walgate

Franco-Soviet space flight

Some hitches

The visit of "spacinaute" Jean-Loup Chrétien to the Salyut-7 space station last month has introduced an unprecedented degree of candour into the reporting of Soviet space flights. Usually, cosmonauts' post-flight press conferences gloss over technical hitches and report as "satisfactorily accomplished" even those missions that were quite obviously curtailed.

The Soviet planners clearly accepted that the presence of a Westerner would entail more candid reporting, and a few days before the launch even admitted that Salyut-7 suffered unexpected attitude changes that had had to be corrected before link-up. Attitude, however, continued to give trouble throughout the flight, both aboard the Soyuz-T transport craft and on Salyut-7 itself.

According to Chrétien, the on-board

computer of Soyuz-T suddenly ceased its automatic plotting just as the craft was approaching the space station at speed. This left Soyuz-T in a rolling motion, from which the crew of three had to effect a manual docking. This, Chrétien later explained on French radio, was not a breakdown — but it came at a critical moment when the crew could not even see the space station.

Orientation problems continued on board Salyut. The French contribution to the programme included high-sensitivity photography of the Earth's atmosphere and the interplanetary environment in the visible and near-infrared range. Another experiment was to photograph very weak light sources, from noctilucous clouds to distant galaxies. Both experiments required both darkness and a high degree of stability.

According to the crew commander, Vladimir Dzhanibekov, both these factors caused difficulty. The cosmonauts repeatedly lost their equipment (cameras, cassettes and so on) in the dark. Worse, flare from the thruster rockets necessary to maintain stability interfered with the lengthy exposures necessary. Dzhanibekov said that such a possibility had never occurred to them before the launch.

Similar problems seem also to have arisen with the various physiological experiments. The crew is said not to have been able to think them through before launch and therefore to have had to "carry out additional training" in space. This seems remarkable — the focus of the medical programme, the French "echography" project, a Doppler measurement of blood-flow in the arteries leading to the brain, was intended to compare cardiovascular function before, during and after the flight. The crew should therefore have been familiar with the apparatus long before take-off.

The apparent lack of preparation may perhaps be due to the international nature of the flight which, contrary to Soviet practice, tied the launch to a pre-announced date.

Vera Rich

Colonel Jean-Loup Chrétien (left) meets President Mitterrand at the Elysée Palace.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

*The "indicated experiments are ALEPH (spokesman Jack Steinberger), OPAL (Aldo Michelini), L3 (Samuel Ting) and DELPHI (Ugo Amaldi). ELECTRA (Roger Cashmore) and LOGIC (Jim Kirkby) have been put on hold.

CORRESPONDENCE

Nuclear dangers

SIR — Nils Ove Bielsten (*Nature* 25 March, p. 284) claims on the basis of "fact" that nuclear power plants are safer than hydroelectric, causing far fewer fatal casualties (a rather dubious claim considering comparative risk analysis), "not a single one occurring to people outside the plants (except possibly in accidents due to construction work traffic)". Yet again the impeccable record of nuclear power is invoked and compared with other sources of energy.

Is it not time to recognize and admit to the dangers of the whole chain of supply of nuclear fuel from mining to its eventual and undecided disposal? Uranium mines where studies have been possible, including some in the United States, have recorded mortality rates between 50 and 100 per cent. Surrounding native American communities also suffer casualties with the additional destruction of their environment and traditional livelihoods, especially painful for people with a traditional respect for the health of the Earth. I believe that it is time for a wider perspective to be taken on large scale projects such as this. Debate over social cost-benefit analyses, environmental impact assessments and technology assessments, which are prone to value judgements, does not negate their usefulness.

The casualties of Britain's earlier plutonium/electricity plants are also becoming apparent and substantial sums have been paid in compensation for workers' deaths.

The economics of nuclear plants are not quite as simple as Mr Bielsten would have us believe either. Left to the electric utilities, as he approves, we find that in the United States no new plants have been ordered for years and several have been cancelled. Plant cost escalations have run up crippling debts for communities and have sent utilities to the federal government for help. In the United Kingdom, it is now suggested that Scottish electricity consumers are paying the price for an unnecessary plant at Torness whose claimed "robust economic case" rests on a possible four-year period when, if coal plants are closed early, it may be useful. Government and independent investigations cast some doubt on even this value to consumers now that costs, need and forecasts are at last being debated.

The writer claims that "nuclear power is the environmentally soundest way of producing the energy that we cannot do without". Yet the people of Sweden have voted not to build any further plants beyond those already started. Their closure is already decided and a transition has been embarked on for a programme of energy efficient development — ending waste to liberate extra available energy. Renewable energy technologies are increasingly the source of new space heating needs and other requirements.

This approach, and energy efficient developments in the United States, demonstrate the value of demand management as a crucial aspect of energy planning.

Finally, Mr Bielsten's reference to possible benefits of mutation in the population is worrying. If it is a serious point then further debate is necessary.

MARTIN S. FODOR

Milton Keynes, UK

Complicating SI

SIR — Like J.A. Nicoll (*Nature* 10 June, p. 450) and probably many others I have considered a simplified notation for physical quantities in terms of SI units.

The appropriate unit is automatically implied by the quantity that must necessarily be specified. Thus the statement that a force is 12 should denote by the absence of units that the SI system is employed and that the force is thus 12 m kg s⁻² or 12 N. Whenever SI is not utilized units must continue to be provided as in 0.237 lb ft min⁻². The convenience of omitting units might thus speed the general adoption of SI.

Prefixes (kilo etc.) are not merely clumsy but they conflict with a principal virtue of SI which is coherence. The nuisance of large or small numbers could of course be lessened by the method commonly practised in computer printouts where 67,000 is written as 6.7 E + 4. However, a simpler notation might be 4;6.7 which specifies the numbers as read from left to right in increasing detail. The semicolon indicates that the number to the left is the power of ten and the period between 6 and the 7 might in time be dropped by convention. Such numbers are subject to rather obvious and simple rules in algebraic procedures.

In this notation the speed of light is 10;2.99776 and the electron rest mass -31;9.109. For the rather unusual case where a physical quantity has a negative value some further symbolism would be required. Perhaps one might adopt the method used by accountants to indicate net loss rather than net profit. In this case the absolute zero temperature on the centigrade scale would be (-2;2.73) C.

Here the unit would have to be provided since the Kelvin scale can also be employed. Amongst other virtues the latter does not have negative values and one would have -∞;0. Which seems a bit odd but could of course also be written conventionally as 0.

HARALD H. ROSSI

Department of Radiology,
College of Physicians and Surgeons
of Columbia University,
New York, USA

German universities

SIR — Thank you for the splendid review on "Science in West Germany", which was published in *Nature* on 27 May. There are, however, some additional points on the present situation of German universities which I would like to highlight.

The statement that government support of science will decrease by 2 per cent in real terms in 1982 is not quite true for the universities. For instance the universities in Nordrhein-Westfalen are facing severe budget cuts which give them only 55 per cent of the 1980 budget. For some institutes in the natural sciences this is less money than they had in 1975, and student numbers have risen by some 50 per cent since; readers of *Nature* will know themselves how much they had to spend on lab-ware and chemicals in the meantime.

Another point is that the measures taken to reform German universities are in fact cutting the numbers of teaching and research personnel at the universities. Only 50 per cent of the young scientists who pursued a

university career and "habilitated" themselves in recent years are eligible for a university job in the new frame law-university structure. The Nordrhein-Westfalen measures to concentrate university studies especially in teacher training, which date from March 1982 and were covered in the *Nature* review, are said to last for at least 5 years in order to transform university jobs into jobs at the Aachen *Klinikum*; at the end of these 5 years the numbers of personnel in the physics and biology departments at Bonn University will have been reduced by 50 per cent.

Since nobody is talking about student number reductions, please add all this up: what you come out with is a catastrophe. Already experimental work for diploma and doctorate theses in biology often is not financed by the universities, but by money from Deutsche Forschungsgemeinschaft (DFG). Although nobody likes to talk about it aloud, DFG funds are also being spent for teaching jobs and so on. Teaching is severely affected: at Bonn the biology department will have to cut lab courses by one third, the chemistry department is seriously thinking about closing down its central chemicals store. All this makes orderly studies and research impossible. The outlook for Germany's universities is very, very grim.

THOMAS SIMON

Bonn, West Germany

Living in the past?

SIR — It would be unfortunate if the pungent style of S. Blinkhorn's review of *The Mismeasures of Man* by S. J. Gould, (*Nature* 8 April, p. 506) should weaken the force of his criticism. Like Blinkhorn I think there is much to object to in this book and, for T. J. M. Schopf (*Nature* 13 May, p. 98), I would like to document a few of its substantive faults.

Although he has references as recent as 1979, Gould ignores major studies and references of the 1970s relevant to the topics he is considering in depth. Consequently, the reader will not learn from Gould's book that most geneticists are not "biological determinists"^{1,2} nor that modern studies of heredity of cognitive abilities do not rely on a single (mystical) measure of intelligence³.

Gould devoted 83 pages to the demolition of nineteenth century claims for an association between cranial capacity and human intelligence and the casual reader undoubtedly will think that he has thoroughly covered this topic. But he omits a critical 1974 review of the subject by Van Valen⁴ which concludes that there probably really is a small positive correlation. Interestingly, Van Valen's conclusion is also ignored by Lewontin (discussing the same subject) in his favourable review⁵ of Gould's book although Van Valen thanks Lewontin for comments (as does Gould).

Gould documents in great detail some blatant errors of the past but, by not coming up to the present, misleadingly implies that it is no different from the past. One would like to know why.

T. EDWARD REED

Departments of Zoology and Anthropology,
University of Toronto, Canada

1. Russell, E. S. *Genetics* 83, s99 (1976).
2. Dobzhansky, Th. *Persp. Biol. Med.* 19, 156 (1976).
3. DeFries, J. C. et al. *Nature* 261, 131 (1976).
4. Van Valen, L. *Am. J. phys. Anthropol.* 40, 417 (1974).
5. Lewontin, R. C. *New York Rev. Books* (22 Oct. 1981).

NEWS AND VIEWS

New branches for old roots

from Barry Cox

THE battle seems now to have been joined in earnest. Waving their branching weapons of attack, some of the cladists are now assaulting even the most fundamental roots of the systems of relationship that the traditional evolutionary morphologists have developed. After the skirmish over the tetrapod-origin outpost, in which Rosen *et al.* (*Bull. Am. Mus. nat. Hist.* **167**, 159; 1981) suggested that lungfishes rather than rhipidistians were closest to tetrapod ancestry, Gardiner has now used cladistic techniques to transform the whole system of tetrapod classification (*Zool. J. Linn. Soc.* **74**, 207; 1982).

Gardiner's method is first to return to the criteria given by nineteenth century anatomists in their analyses of relationship of the major tetrapod groups. Thus he concludes that "the denial of any immediate relationship between the mammals and the other great homotherm (*sic*) group, the birds, was based on no more than three characters; the form of the heart, aortic arches and occipital condyle". These three characters he swiftly demolishes (in about a page of text), and immediately moves on to list a number of characteristics of homoiothermy and of the central nervous system that appear to link mammals and birds as sister-groups. He relates to these two, in order of increasing distance, the crocodiles, chelonians and lepidosaurs. In all these discussions, palaeontology is ignored and only the characteristics of living animals are used. Into the resulting classification, Gardiner then fits a few fossil groups, such as the pterosaurs (a sister-group of birds), dinosaurs in Owen's original sense (sisters to crocodiles), the cynodonts (far divorced from the mammals), the dicynodonts (sister-group to the chelonians) and the pelycosaurs (completely separated from the previous two groups). The results are so different from those that also incorporate palaeontological data that a headlong collision is inevitable, for there can be no minor shifting of viewpoints or emphasis that could lead to a reconciliation or synthesis.

Gardiner's methods can be illustrated particularly well in his treatment of the Synapsida, the splitting of which into three separate groups (pelycosaurs, dicynodonts and cynodonts) is, to a palaeontologist, one of the most revolutionary aspects of his work. Basically, he rejects previously established relationships if he can quote other workers to the contrary and can show

that some members of any other group also share some of the features previously thought to support the accepted relationship. So "as long ago as 1926 Berg wrote that the mutual similarities between mammals and (advanced synapsids) were probably due to convergence" and "despite the fact that fenestrae might have arisen independently (as they have done within the Chelonina . . . and Amphibia . . .), today's unquestioned doctrine is that the Synapsida include the mammal's closest relatives".

Gardiner also underestimates the modern palaeontological evidence by restricting himself to out-dated authors; it is really true of the pelycosaurs that "Their claim to Synapsid relationship apparently rests on Cope's (1884) insistence on their closeness to mammals and the possession of a lateral temporal fenestra"? Reisz, himself a cladist, states that "The pelycosaurs, together with their descendants, form a natural group and possess such derived characters as broad anteriorly tilted occipital plate . . . reduced post-temporal fenestrae, single median post-parietal, and septomaxilla composed of a broad base and a dorsal process" (*Spec. Vol. Syst. Ass.* **15**, 553; 1980). The relationships of pelycosaurs to both dicynodonts and cynodonts (and also to other therapsid and theriodont groups) is also suggested by the fact that all these groups share with sphenacodontid pelycosaurs the synapomorphic feature of a reflected lamina on the angular bone of the lower jaw. All, except the turtle-beaked dicynodonts, also share in having a heterodont dentition.

The cynodonts, like some other advanced synapsids, show a progressive reduction of the post-canine dentition and of the post-dentary bones of the lower jaw, with the mammalian type of squamosal-dentary jaw articulation having evolved in a few forms, and various approaches to the mammalian condition in the secondary palate, dentition, limbs and girdles. Gardiner nevertheless dismisses any relationship between cynodonts and mammals because an examination of the British Museum (Natural History) specimens of the cynodont *Thrinaxodon* "failed to

reveal a single mammalian synapomorphy" and, for example, showed that *Thrinaxodon* was not particularly mammalian in its tooth-replacement and braincase. But it is such later, mid-late Triassic cynodonts as *Probainognathus* and the trithelodonts that are thought to be close to the origin of mammals, and these do show far more mammalian conditions of the teeth and braincase.

It seems perverse to reject these complex similarities merely because our extensive knowledge of the osteology of both recent and fossil tetrapods reveals isolated instances where temporal fenestrae have evolved independently (and, in any case, the chelonian openings that Gardiner quotes are emarginations, not fenestrae). We have no such systematic knowledge of the soft-part anatomy, physiology or biochemistry that make up the 17 characters that Gardiner lists to support his contention that birds and mammals are closely related sister-groups. As Szalay (*Am. Zool.* **21**, 37; 1981) has commented: "I believe that at the very heart of the post-Hennigian cladistic approach to phylogeny reconstruction lies the implicit neglect for biological and functional scrutiny of the characters employed. As a result of this attitude the postulated homologies chosen (the units on which phylogenies are based) are not biologically investigated and therefore the chances of discovering whether these features are unique or parallel or convergent acquisitions, or whether they represent a pattern commonly attained by distant lineages because of epigenetic factors, or because they are rooted in a common heritage, weakens the very foundations of such analyses. In other words, character analysis commonly practised in cladistics is really a semantically masqued enumeration of characters without attempts to analyse and weight them".

Gardiner's suggestions are so completely at odds with the palaeontological record that some of Panchen's observations, in an incisive criticism of the cladistic method published in the same volume as Gardiner's work (*Zool. J. Linn. Soc.* **74**, 207; 1982), seem particularly relevant. "The first obvious fault of cladism in general is perhaps a temperamental and, one may hope, a temporary one. This is the temptation to produce new and revolutionary groupings of organisms, or to resurrect old and discredited ones, just to demonstrate that cladism is getting results." □

Barry Cox is a Professor in the Department of Zoology, King's College, University of London, London WC2.

Chaos, randomness and dimension

from T. Geisel

CHAOS and randomness are ubiquitous. In many scientific disciplines, random time series are obtained from an experiment and analysed into their power spectra. Often enough, the irregular motion of a system will have been produced by an external noise source. On the other hand, irregular motion and irregular time series may also be self-generated by a purely deterministic system. In the latter situation the term 'chaos' is used and, in a state space description, its dynamics are often governed by strange-looking topological objects called 'strange attractors'. As the scientific community is becoming aware of this phenomenon, observations of irregular behaviour that previously might have been ignored are attracting attention. The following questions then arise: given only the irregular time series, is it possible to tell whether it is due to chaos in the above sense, that is, does it have intrinsic properties which allow such a decision? Is it even possible to reconstruct the strange attractor, the ID card of the system, from the time series? These questions have been addressed in several recent papers which are reviewed in this article.

When irregularity is due to thermal fluctuations in physical systems, a large

number (10^{23}) of particles and degrees of freedom are involved, which make the measurements look noisy. In brownian motion, for example, the irregular motion of a heavy test particle only reflects the irregular dynamics of 10^{23} molecules in the surrounding liquid. The terms 'randomness' and 'noise' describe such situations. 'Chaos', in contrast, is reserved for an irregularity that is self-generated by a deterministic system with few degrees of freedom in the absence of external fluctuating forces. Simple examples are driven anharmonic oscillators¹, where the phase space is three dimensional, consisting of position x , velocity v and the phase ψ of the driving force. In an undriven damped harmonic oscillator, the motion would asymptotically settle down at $x=0$, $v=0$. This point is called an attractor because any initial condition x_0 , v_0 in a given area of phase space would lead to this point. The dimension is thereby shrinking from two to zero. For chaotic systems which are dissipative, attractors called strange attractors^{2,3} may occur. These are asymptotic limit sets in phase space which also have lower dimension than the phase space and which have an additional property: the motion on the attractor is characterized by an exponential divergence of neighbouring states in some direction. This sensitive dependence on initial conditions is the cause for their strange

appearance. It also practically prevents long-term predictions since the initial conditions are usually not known exactly. The sensitive dependence and the degree of unpredictability are measured by a variable called a Lyapunov exponent.

Strange attractors may not only occur in systems with few degrees of freedom but also in systems where, *a priori*, the dimension of phase space is high or infinite. An example is provided by turbulence in fluids and other systems that can be described by partial differential equations. Due to the presence of dissipation (damping) the motion in multidimensional phase space may, with time, settle down on a strange attractor which acts like a low-dimensional trap. The motion can then be described by a few variables in phase space. If the topology of the strange attractor is known, the dynamics of the entire system is known. This explains the dominating role of low-dimensional strange attractors for the chaotic dynamics of systems that may have infinite dimensions. Due to the sensitive dependence on initial conditions, these attractors have a complex structure similar to a Cantor set. The fractal dimension^{4,5} can be used to characterize their spatial organization. In contrast to the topological dimension, this number may be non-integral, reflecting the topological complexity.

When an irregular time series is observed in an experiment one may ask whether the irregularity is due to chaos or to randomness as distinguished above. When little is known about the system, one can only hope that the time series possesses intrinsic properties which make this distinction possible. Power spectral analysis, which is usually applied, is of little help but other techniques have recently been proposed. They can be used to determine the dimension of the underlying dynamics and the fractal dimension of strange attractors. It is even possible to reconstruct the strange attractors.

One procedure has been proposed by the Santa Cruz Dynamical Systems Collective (J.P. Crutchfield, J.D. Farmer, N.H. Packard and R.S. Shaw). They propose to use the observed signal $x(t)$ and its successive derivatives $\dot{x}(t)$, $\ddot{x}(t)$, ... as variables spanning the subspace of phase space to be reconstructed⁶. The procedure has been applied successfully to a dynamical system due to O.E. Rössler. It cannot be applied when the time series is not detailed enough to determine the derivatives. Another procedure, suggested by D. Ruelle (personal communication), consists of using the signal $x(t)$ and delayed signals $x(t-\tau)$, $x(t-2\tau)$, ... to span the phase space reconstruction. F. Takens⁷ has given a formal justification of these procedures in terms of the Whitney embedding theorem.

Roughly speaking, the dimension is determined if, with increasing dimension, the objects produced by the motion start

T. Geisel is at the Institut für Theoretische Physik, Universität Regensburg, D-8400 Regensburg, FRG.

100 years ago

THE NEW AFRICAN EXPEDITION

It is now understood to be quite settled that a new African exploring expedition will start next year. The Royal Geographical Society have, as might have been expected, taken the opportunity of Mr. Joseph Thomson's return from the completion of his engagement to the Sultan of Zanzibar to obtain his services as leader, and it is certain that no better selection could have been made.

Mr. Thomson will leave England in the Spring of 1883, and proceed to Zanzibar to organise the expedition. From Mombas, a port on the East African coast, to the north of Zanzibar, he will direct his course straight to Kilimandjaro, and do his best to explore the snowy ranges of this celebrated mountain, which but one European has as yet ever reached. Passing across the waterparting he will then descend through an entirely unknown country to the eastern shore of Lake Victoria Nyanza, and return to the coast by a more northern route, in the course of which it is hoped he may be able to visit Lake Baringo and Mount Kenia — another peak known to run far above the snow-level, but concerning which further details would be very desirable.

As a mere geographical expedition it will be

thus seen that the proposed route will be one of great interest, embracing, as it does, the transit through much utterly unknown country, and the exploration of two mysterious snow-crowned mountains, which, according to the usual view of the conformation of the African Continent, appear to be quite out of place in the districts in which they are situated. But still more interesting problems will be solved, if steps are taken to investigate the unknown fauna and flora of Kilimandjaro and Kenia.

GEOGRAPHICAL NOTES

Lieut. Giraud has sailed from Marseilles for Zanzibar, as leader of a French expedition which proposes to take up African exploration where Livingstone laid it down with his life on the south shore of Lake Bangweolo. Lieut. Giraud proposes to go either direct west to Lake Tanganyika, or, more probably, by the north shore of Lake Nyassa, to the Chambeze River. This he will follow to its outlet in Lake Bangweolo, which he proposes to circumnavigate he will then attempt, in canoes, to sail down the Luababa-Congo, to its mouth in the Atlantic Ocean. This is an ambitious programme; and every one interested in African exploration will wish the expedition complete success.

From *Nature* 26, 304 & 309, 27 July, 1881.

making sense. More systematic techniques, however, are now available. The Dynamical Systems Collective uses linear regression locally to fit the data to points, planes or higher-dimensional hyperplanes⁵. The goodness of the fit is determined by χ^2 , the sum of the squared deviations divided by the dimension. When the dimension closest to the fractal dimension is reached, χ^2 drops sharply. In the example of the Rossler system, a drop by a factor of one million was found for the best fit. Once the strange attractor has been reconstructed, the fractal dimension may be determined according to its definition⁵. Techniques to determine the fractal dimension and another measure of chaos, the topological entropy, have been described by F. Takens⁷. The relationship between the fractal dimension and the Lyapunov exponents has been investigated by P. Frederickson *et al.*⁸.

Procedures which are able to distinguish between chaos and randomness are also suggested in an article by J. Guckenheimer published in this issue of *Nature* (p.358). These procedures are based on the assumption that randomness means unpredictability of the short-term evolution of a system. Chaos, on the other hand, is characterized by an exponential divergence of neighbouring states in phase space. This brings about only a long-term unpredictability of the system. Consequently procedures are constructed that verify exponential divergence of neighbouring states in the reconstructed phase space. Exponential divergence has also been investigated in terms of the Lyapunov exponent by R.S. Shaw for a

chemical reaction system⁹.

The methods outlined here have been applied successfully in real experimental situations. Strange attractors have been reconstructed for chaotic chemical reactions by J.C. Roux *et al.*^{10,11} by using the time series of the concentration of a single chemical species — the phase space may in principle consist of at least the concentrations of all chemical species involved (about 25). D. Farmer *et al.*¹² have recently used a phase space reconstruction for a baroclinic flow experiment. In addition, they have been able to determine quasi-one-dimensional Poincaré return maps. R.S. Shaw *et al.*¹³ have studied the irregular dripping of a tap and derived one- and two-dimensional return maps and topological and metric entropies from their measurements. □

1. Huberman, B.A. & Crutchfield, J.P. *Phys. Rev. Lett.* **43**, 1743 (1979).
2. For an introductory article see, for example, Ruelle, D. *La Recherche* **11**, Feb. 1980, p.132; English translation in *Math. Intelligencer* **2**, 126 (1980).
3. Ruelle, D. & Takens, F. *Comm. Math. Phys.* **20**, 167 (1971); **23**, 343 (1971).
4. Mandelbrot, B.B. *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).
5. Froehling, H., Crutchfield, J.P., Farmer, D., Packard, N.H. & Shaw, R. *Physica* **3D**, 605 (1981).
6. Packard, N.H., Crutchfield, J.P., Farmer, J.D. & Shaw, R.S. *Phys. Rev. Lett.* **45**, 712 (1980).
7. Takens, F. in *Dynamical Systems and Turbulence* (eds Rand, D.A. & Young, L.S.) *Lect. Notes Math.* **898**, 366 (1981).
8. Frederickson, P., Kaplan, J.L., Yorke, E.D. & Yorke, J.A. *J. different Equat.* (in the press).
9. Shaw, R.S. as quoted in Roux, J.C. & Swinney, H.L. in *Nonlinear Phenomena in Chemical Dynamics* (eds Vidal, C. & Pacault, A.) **38** (Springer, Berlin, 1981).
10. Roux, J.C., Rossi, A., Bachelart, S. & Vidal, C. *Phys. Lett.* **77A**, 391 (1980); *Physica* **2D**, 395 (1981).
11. Turner, J.S., Roux, J.C., McCormick, W.D. & Swinney, H.L. *Phys. Lett.* **85A**, 9 (1981).
12. Farmer, D., Hart, J. & Weidman, P. (in preparation).
13. Shaw, R.S. *et al.* (in preparation).

mitochondria was examined by preparing tRNA from well-washed, isolated organelles, digested and analysed by HPLC. What was striking was that the purified mitochondrial tRNAs from wild type contained both methylated nucleosides, but m²G and m⁵U were specifically lacking in the mitochondrial tRNA from the *trm1* and *trm2* mutants respectively. Thus, the phenotype of these mutants is the same at the level of whole-cell (primarily cytoplasmic) tRNA and mitochondrial tRNA. On the face of it, it appears that for both modifications, the same (or similarly regulated) enzyme activities function inside and outside the mitochondria. (One assumes that the tRNAs do not leave the matrix to be modified and then return.)

It is dangerous to jump from the fact that a mutation affects an enzyme activity inside and outside the mitochondria to the conclusion that the mutation is in a gene coding both enzymes, without examining the enzymes themselves. Studies by Gross and co-workers on the *LEU5* locus in *Neurospora crassa* have shown that a mutation there leads to the production of an altered cytoplasmic leucyl-tRNA synthetase and to the virtual lack of mitochondrial leucyl-tRNA synthetase. Both phenotypes co-revert, indicating that they are the result of a single mutation. Furthermore, the genetic determinant for a naturally occurring electrophoretic variant of the mitochondrial enzyme also maps to the *LEU5* region, indicating that structural gene information for both enzymes is in the same place (Beauchamp *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 1172; 1977). However, they could find no immunological cross-reactivity between the cytoplasmic and mitochondrial enzymes leading them to propose two tightly linked or overlapping genes for these two enzymes. Since the lack of immunological cross-reactivity may not be the most reliable indication that two proteins are unrelated, one should not yet rule out the possibility that a single gene codes for two, perhaps differently modified, leucyl-tRNA synthetases in *N. crassa*. However, in the case of the *trm1* and *trm2* mutations of yeast, the possibility of distinct but tightly linked and co-regulated genes for tRNA methylases should probably not be dismissed either.

A single gene, with two possible modes of transcription and translation, could clearly code for two similar proteins that end up on opposite sides of a membrane, as has been shown by analysis of a yeast gene that codes for both intracellular and secreted invertase (Perlman and Halvorson *Cell* **25**, 525; 1981; Carlson and Botstein *Cell* **28**, 145; 1982). It seems quite possible that such a mechanism also plays a part in specifying some proteins inside and outside the mitochondrial matrix. Clearly, an examination of the *TRM1* and *TRM2* genes of yeast and the *LEU5* gene of *N. crassa*, as well as the proteins they encode, will prove very enlightening. □

Do mitochondria share enzymes with the rest of the cell?

from Thomas Fox

In a rough way, the matrix of mitochondria can be thought of as a cell cytoplasm containing a chromosome, a transcription and translation apparatus, and a number of metabolically important soluble enzymes. The peculiar thing, of course, is that much of this — for reasons that are unknown — apparently duplicates functions present in the real cell cytoplasm and the nucleus. It has seemed, however, that in no case does the same protein species function both inside and outside the mitochondrial matrix. Even though virtually all of the proteins present in the matrix are coded by nuclear genes, the matrix proteins studied so far have been found to be biochemically distinguishable

from their functional analogues outside the mitochondria. A new report from Hopper *et al. (Cell* **28**, 543; 1982) now provides evidence that this may not always be the case. Work on two yeast mutants that affect tRNA methylation suggests that these functions are indeed shared by mitochondria and their gracious hosts.

The nuclear mutations, *trm1* and *trm2*, cause the absence of tRNA (guanosine-N², N²) dimethyltransferase and tRNA (uridine-5)methyltransferase activities respectively. This was clearly demonstrated by examining HPLC elution profiles of ribonucleosides derived from digests of whole-cell tRNA preparations: m²G is specifically absent from the tRNA of the *trm1* mutant, while m⁵U is specifically absent from the tRNA of the *trm2* mutant. The modification of those tRNAs encoded and transcribed within

Thomas Fox is Assistant Professor of Genetics, Division of Biological Sciences, Cornell University, New York 14853.

Bone in the cartilaginous fishes

from Brian Hall

THE recent documentation by Peignoux-Deville *et al.*^{1,2} of bone in the vertebrae of the backbone of the dogfish, *Scyliorhinus canicula*, a member of the cartilaginous fishes (Chondrichthyes), has important implications for biology because sharks and rays are grouped together as cartilaginous fishes on the basis of the presence of a cartilaginous skeleton in the adult and the absence of bone from any stage of the life cycle. In other vertebrates, such as man or the bony fishes, most of the cartilage of the embryo is replaced by bone in the adult.

Cartilage and bone are the two major skeletal tissues. The long debate in evolutionary circles over whether cartilage or bone arose first during vertebrate evolution resulted in classifications of the vertebrates that had cartilaginous fishes either on the main evolutionary line as the ancestors of all other fishes and amphibians, if you believed that bone arose late in vertebrate evolution, or as a side branch, if you believed that bone arose early³.

Peignoux-Deville *et al.* examined vertebrae of the dogfish using X-ray analysis, labelling with tetracycline (a calcium-seeking fluorescent drug) and transmission electron microscopy. The upper halves (neural arches) of the vertebrae were found to consist of large-celled cartilage overlain by a layer of lamellar bone containing bone-forming cells (osteocytes). The bone was calcified, as confirmed by the uptake of tetracycline and by the presence of calcium phosphate (apatite) crystals. Average mineralization (1g mineral per cm³ of bone) was the same in the cartilaginous and bony portions of the vertebrae. The association of the bone with the underlying large-celled cartilage suggests that the cartilage may have induced the bone to form in the same way as occurs in higher vertebrates⁴. Mutants are known in which cartilage fails to become large-celled (for example, *talpid* in the chick, *brachypod* in the mouse), with the consequence that bone fails to form around the cartilage. Cartilaginous fishes appear to use the same developmental mechanism to form bone. The cartilage of higher vertebrates also contains canals which both provide cells for centres of bone formation deep within cartilage and allow nutrients to reach the deeper recesses of cartilage, a tissue which lacks blood vessels. Hoenig and Walsh⁵ examined vertebrae from 16 species of sharks and found that all had canals containing blood vessels and immature cartilage cells. They ruled out any bone-forming (osteogenic) function for these canals . . . "the

osteogenic function cannot apply to Elasmobranch vertebrae", following instead the traditional view that sharks lack bone. A closer examination of these vertebrae is clearly now required.

Previous studies had hinted at the presence of bone in the endoskeletons of sharks and definitely showed it to be present in the hard denticles which are embedded in the skin and at the base of the teeth⁶⁻¹¹. The X-ray diffraction pattern of shark vertebrae is also identical to that obtained from bone¹², and this and the presence of the canals should encourage a closer examination of shark vertebrae. *Ornithoprion hertwigi*, a fossil shark from the Permian period, had a thick layer of bone around its lower jaw¹³. The presence of bone in both the jaw and the vertebrae is of especial interest as much of the head skeleton is derived from the neural crest^{4,14}, while vertebrae are mesodermal in origin. Sharks have retained bone-forming potential in cells with both of these embryological origins. Bone formation in cells on the surface of the vertebrae is presumably activated by association with underlying large-celled cartilage while cranial neural crest-derived cells are presumably activated by interaction with epithelia, as happens in other vertebrates¹⁴.

The existence of bone in these sharks explains the previously paradoxical presence of the hormone calcitonin, which acts on bone to regulate calcium

metabolism. The presence of bone also highlights the channelling of thinking which follows the classification of a group of animal on the basis of the presence of one, albeit major, feature and the lack of another. Because the Chondrichthyes are classified as having a cartilaginous skeleton and lacking bone, we automatically think that they are incapable of forming bone. This led to the spurious classifications of vertebrates mentioned earlier, to the conclusion that no feature (canals, calcitonin) could have anything to do with bone or bone formation, that the cartilaginous fishes have 'lost out' in evolution, not having the potential for growth, mineral metabolism and structural support possessed by their bony cousins, and that any suggestion of the presence of bone must be spurious for cartilaginous fishes lack bone. The simple observation of the presence of a tissue in one region of one species opens up many roads for future travel. □

1. Peignoux-Deville, J. *et al.* C.R. hebdomadaire Séances Acad. Sci., Paris 292, 73 (1981).
2. Peignoux-Deville, J. *et al.* Cell Tissue Res. 222, 605 (1982).
3. Hall, B.K. Am. Zool. 15, 329 (1975).
4. Hall, B.K. Developmental and Cellular Skeletal Biology (Academic, New York, 1978).
5. Hoenig, J.M. & Walsh, A.H. Can. J. Zool. 60, 483 (1982).
6. Kemp, W.E. & Westrin, S.K. J. Morph. 160, 75 (1979).
7. Moss, M.L. Am. Zool. 17, 335 (1977).
8. Holmgren, N. Acta zool. 23, 129 (1942).
9. Moss, M.L. Acta anat. 77, 161 (1970).
10. Peyer, B. Comparative Odontology (University of Chicago Press, 1968).
11. Miles, R.S. & Moy-Thomas, J.A. Palaeozoic Fishes 2nd edn (Chapman & Hall, London, 1971).
12. Gilbert, P.W. *et al.* Sharks, Skates and Rays, 37 (The Johns Hopkins Press, Baltimore, 1967).
13. Zangerl, R. Fieldiana Geol. 16, 1 (1966).
14. Pratt, R.L. & Christensen, R.L. Current Research Trends in Prenatal Craniofacial Development (Elsevier, New York, 1980).

Hipparcos: space-age astrometry

from A.M. Cruise

THE measurement and cataloguing of stellar positions is one of the oldest branches of science and the results, gathered over the centuries, have been crucial in the formulation of our present concept of the scale and structure of the Universe. But astrometry is hardwork. To achieve a precision very much better than the arc second resolution of ground-based optical telescopes in a homogeneous manner over the whole sky seems impossible, with present instruments at least. It is no wonder then that astrometry has fallen from the front ranks of astronomical science, and spectroscopy and theoretical astrophysics have gained favour. Nevertheless, astrometry has some simple but profound truths to tell that are fundamental to the development of all branches of astronomy. By choosing the space astrometry mission 'Hipparcos', scheduled for launch in late 1986, the European Space Agency will transform

many branches of astronomy and astrophysics.

The astrometric observations made from Hipparcos are hoped to have a precision of one or two milliarc seconds — one hundred to a thousand times better than is currently available for most stars. Such accuracy will allow both radio VLBI measurements and optical results to be specified in a consistent coordinate system.

Increased precision is not the only attribute of the Hipparcos catalogue. The instruments scanning sequence will allow observation of the whole celestial sphere during the two and a half year mission. About 1000,000 stars will be measured, including all stars down to ninth magnitude, and careful selection will ensure that there are always two or three

Brian Hall is Professor and Chairman of Biology, Dalhousie University, Halifax, and a Nuffield Foundation Fellow and Visiting Fellow, University of Southampton.

A.M. Cruise is at the Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT.

Hipparcos stars in each square degree of sky. A convenient, homogeneous coordinate grid of high precision that covers the whole sky will thus be created. In addition, the Space Telescope will be able to determine the position of some extragalactic objects with respect to Hipparcos stars and so link the new coordinate system to an inertial frame.

The increase in precision gained from Hipparcos will allow the absolute determination, by trigonometric parallax, of distances 10–20 times greater than at present. It is the absolute nature of the method which is of overwhelming importance. All the overlapping cosmic distance scales which span the observable Universe ultimately depend on measurements of the distance to the Hyades cluster which is just close enough to show measurable angular motion in ground based telescopes. Calibration of the absolute luminosity of various stellar distance indicators in the Hyades provides the means by which apparent luminosities of very distant objects can be used to determine their distance.

Not only will Hipparcos improve on the absolute distance measurement for the Hyades and so reduce the statistical error on all cosmic distance scales but more important, the absolute distance of five clusters other than the Hyades will be measured. The calibration of distance indicators can then be scrutinized for dependences on metallicity and other factors that might vary from cluster to cluster. Hipparcos will also increase the number of stars accessible to direct distance measurement by a factor of about 1,000 and absolute luminosity measurements over a much wider range of stellar temperature will be made possible.

The accuracy of measurement will allow changes in stellar positions to be observed during the mission. This proper motion caused by the stars spatial velocity with respect to the Sun has been measured in the past for a relatively small number of stars by comparing positions measured at widely separated times. The high precision of Hipparcos will, however, be offset by the shortness of the two and a half year observation period so the accuracy of proper motion measurements will not be much increased — the really dramatic improvement will be in the large numbers of stars observed.

The performance of the Hipparcos instrument is closely coupled to the operation of the satellite and the design of both is being refined. The two main difficulties, which should be overcome by the concept first proposed by Lacroute, are the achievement of the necessary precision and the maintenance of the absolute calibration of the angular measurements. A modestly sized telescope will view a beam combiner so that two stars separated on the sky by about 70 degrees will be superimposed in the focal plane on a modulating grid that measures their relative separation.

As the satellite rotates the two fields of view will scan along a great circle. By demanding that a closed chain of star separation measurements must sum to 360 degrees angular calibration can be accomplished with the desired frequency. When many intersecting great circle scans are combined the two coordinate stellar position can be found with an ultimate accuracy of one or two milliarc seconds.

There is already a great deal of scientific activity, including colloquia^{1,2} and published papers^{3–5}, supporting the Hipparcos project. ESA has selected two separate consortia to perform the data analysis task and has issued a call for proposals to help define the list of stars to be observed. The deadline of 1 October

1982 gives some hint of how much work is required to put the observing list into the necessary form for the operation of the satellite.

Astrometry from space will still be hard work but at least the astrometric measurements from Hipparcos will stand more clearly above the noise and so attract the attention they deserve from astronomers in general. □

1. Barvieri, C. & Bernacca, P.L. *Colloq. European Satellite Astronomy*, University of Padova (1978).
2. ESA SP-177 *The Scientific Aspects of the Hipparcos Space Astrometry Mission* (ed. Perryman, M.A.C. & Guyenne, T.D.).
3. Hoyer, P. *et al. Astr. Astrophys.* **101**, 228 (1981).
4. Murray, C.A. *Mon. Not. R. astr. Soc.* **195**, 639 (1981).
5. Froeschle, M. & Mignard, F. *Appl. Optics* **20**, no. 18, 3521 (1981).

Monoclonal antibodies to human malignant melanoma

from Ralph A. Reisfeld

A RECENT international workshop on monoclonal antibodies to human melanoma-associated antigens* demonstrated rapid progress in the analysis of 30 monoclonal antibodies that had been produced and exchanged by 10 laboratories. There was general agreement that 24 antibodies defined 11 antigens that are expressed on melanoma cells to a considerably greater extent than on other normal or neoplastic cells and tissues. These antigens may be the products of amplified and/or altered normal genes and could provide a set of model targets for adjunct immunotherapy, as well as sensitive molecular probes for studies on the biology of malignant melanoma and the mechanism of neoplastic transformation.

Particularly striking were the consistency among reaction patterns of most antibodies, even though different target cells and techniques were used, and the variety of the antigens identified, even by using the same melanoma cell lines as immunogens. Although most monoclonal antibodies were elicited from melanoma lines in long-term tissue culture, they reacted in binding indirect immunofluorescence and immunoperoxidase assays with surgically removed, frozen tissue sections and, in some cases, with melanoma cell lines at early passage levels

— suggesting few tissue culture artefacts and little antigenic drift. Only relatively few antibodies reacted with formalin-fixed, paraffin-embedded tissue sections, suggesting that most determinants recognized by the monoclonal antibodies are irreversibly denatured by this treatment.

Structural studies of the 11 selected antigens indicated 5 of them to be single-chain glycoproteins with molecular weights (M_r) of 70,000–100,000, one a single-chain glycoprotein with M_r 184,000 and two multiple-chain glycoprotein molecules with M_r of 116,000–130,000. One of the antigens is a sialoganglioside (G_{D3}) and two were identified as extracellular matrix proteins, that is, chondroitin sulphate proteoglycans consisting of a glycoprotein chain (M_r 250–280,000) and a series of components with M_r > 780,000. Antibodies to two of the antigens, p97 (gp95), a glycoprotein of M_r 97,000 which is structurally related to transferrin, and the proteoglycan molecules, have been identified which can react with distinct epitopes on the same or very similar molecules. Interestingly, all the 24 monoclonal antibodies selected react with cell membrane or matrix antigens. This may reflect a relatively low immunogenicity of cytoplasmic antigens and/or the choice of the procedures used in screening the hybridomas.

Several general conclusions were reached by the 10 participating laboratories. The antibodies analysed detect tumour-associated antigens but not tumour-specific antigens in the strict sense

*The workshop on 'Monoclonal Antibodies to Human Melanoma Antigens: II' held on 26–27 April 1982 at Bethesda, Maryland was organized by the Immunology Program, Cancer Biology Branch, Division of Cancer Biology and Diagnosis, National Cancer Institute. The participating laboratories were those of J.C. Bystryn, New York University School of Medicine; S. Ferrone, Columbia University, New York; K.E. and I. Hellstrom, Fred Hutchinson Cancer Center, Seattle; H. Koprowski, Wistar Institute, Philadelphia; J.P. Mach and S. Carrel, Ludwig Institute for Cancer Research, Lausanne; D.L. Morton, University of California, Los Angeles; R.A. Reisfeld, Scripps Clinic and Research Foundation, La Jolla; G. Riethmüller, University of Munich; and H.F. Seigler, Duke University Medical Center, Durham. The proceedings of the workshop will be published in *Hybridoma* **2**, No. 4, 1982.

Ralph A. Reisfeld is at Scripps Clinic and Research Foundation, La Jolla, California 92037.

of the word; those selected are, however, sufficiently closely associated with melanomas that they should be further characterized and serious consideration given to their potential as *in vitro* and *in vivo* markers for diagnosis and as targets for immunotherapy. Simple screening of monoclonal antibodies by binding assays with cultured tumour cell lines is not sufficient to evaluate critically their specificity, and screening of a large variety

of both normal and tumour tissues by indirect immunofluorescence or immunoperoxidase techniques appears necessary to make any meaningful comparisons. Quantification of antigen expression on normal and tumour cells is definitely necessary, either by saturation binding with Scatchard plot analyses or by quantitative binding assays with radio-labelled antibody or by analysis with the fluorescence-activated cell sorter. □

leaves in a canopy (P. Horton, Sheffield University).

The majority of temperature crops, particularly cereals, are exposed to suboptimal temperatures for most of their lives. Dramatic differences in temperature response occur between cold-tolerant and cold-intolerant plants. Even at the level of the chloroplast the partitioning of assimilate may be considerably influenced by temperature as transport generally outstrips production at low temperatures; the autocatalytic Benson-Calvin Cycle does not obey Arrhenius' law (Heber, Würzburg University). Low temperatures cause phase changes in the lipids of the thylakoid membranes and the resulting reduction of the fluidity of the thylakoids reduces the mobility of the electron carriers between photosystems I and II, for example by diminishing the diffusion rate of reduced plastoquinone (D. Chapman, Imperial College, London). This may well be the most temperature-dependent step of electron transport.

Ribulose biphosphate carboxylase/oxygenase is the point of entry of carbon into the Benson-Calvin cycle and that of oxygen into the photorespiratory pathway. There is a broad correlation between ribulose biphosphate carboxylase activity extractable from leaves and light-saturated rates of photosynthesis in plants taken from a wide range of growth conditions and genotypes. Ribulose biphosphate carboxylase appears to limit photosynthesis only in the sense that the activity of the enzyme is close to that of the flux through the Benson-Calvin cycle. Plants synthesize only as much of this enzyme as they need (A. Keys, Rothamsted Experimental Station). This raises the question of whether a limitation imposed by this enzyme is a phenomenon perceived by the manipulator rather than the plant.

The purpose of photorespiration continues to be hotly debated. Selective inhibition of the oxygenase activity is improbable although some species have carboxylases with a significantly higher affinity for CO₂ than others (S. Gutteridge, Rothamsted Experimental Station). Other enzymes of the Benson-Calvin cycle, such as fructose biphosphatase, show changes in activity in response to the redox state of ferredoxin, although the changes in fructose biphosphatase activity do not have a role in regulation of the photosynthetic rate in leaves illuminated with saturating light.

No final answer as to 'What limits photosynthesis' arose from the meeting. The most popular candidate, proposed by O. Björkman (Carnegie Institute of Washington), was God. The question deserves future consideration and research if we are to secure increased plant productivity. □

Christine Foyer, Richard Leegood and David Walker are in the ARC Research Group on Photosynthesis, Department of Botany, Sheffield S10 2TN.

What limits photosynthesis?

from Christine Foyer, Richard Leegood and David Walker

ENHANCED plant productivity is one means by which world food and energy shortages may be alleviated. For this reason research into the photosynthetic process is increasingly favoured both in the UK and overseas. However, the crop physiologist is quick to remind us that the connection between photosynthesis and productivity is more apparent than real, in the sense that a correlation between short-term measurement of photosynthetic rate and crop yield is not easily demonstrated. Certainly the gulf between the plant scientist and the farmer is considerable. Even the highly valuable contributions of the plant breeder have been largely confined to factors such as disease resistance and the redistribution of assimilates within the plant. We have all lamented photorespiratory loss, worried over the inadequacies of ribulose biphosphate carboxylase/oxygenase and wondered if we can turn, for solace, to C₄. Perhaps, however, if we pause for thought, there is no reason to be unduly pessimistic about the future of photosynthesis in its relation to agriculture.

Plant growth is the best measure of net photosynthesis that we have and the limits to which the useful redistribution of assimilates can be achieved are fast being reached. To increase plant productivity by securing more vigorous growth through more effective photosynthesis we need to determine which existing limitation might be removed. A discussion meeting* was organized by the Agricultural Research Council to identify and characterize the reactions which limit the photosynthetic process in various environmental and developmental conditions, and to determine how basic research into photosynthesis may be directed towards increasing plant productivity. A workshop on the methodology and applications of the measurement of chlorophyll fluorescence preceded the meeting. This technique is now regarded as a valuable non-intrusive probe for photosynthesis *in vivo* providing, in the words of J. Barber

(Imperial College, London), a 'stethoscope' for plants.

Many workers in photosynthesis disagree about the basis on which the rate of photosynthesis should be expressed, and doubt the validity of the present methods, especially since a single factor is unlikely to be constant. All quantitative means for expressing photosynthetic rate in current use (for example, ground area, fresh weight) carry inescapable disadvantages. In spite of this, chlorophyll is likely to remain the universal basis for expressing photosynthetic rate at the cellular level simply because it is easily measured (J. Ludwig, GCRI, Littlehampton). The use of the term 'photosynthetic efficacy', to relate capacity to performance (that is, an assessment of the actual performance of a given crop as a function of the potential performance), could help to determine the degree to which photosynthetic capacity is realized in a given circumstance (D. A. Walker, Sheffield University).

Although the basic climatic limitation to crop photosynthesis is the seasonal input of light energy, the degree to which this energy can be used by the plant is modified by local climatic factors, such as temperature, water and the requirements of particular farming systems (J. P. Cooper, WPBI, Aberystwyth). The rate of photosynthesis reaches a maximum at or just before full leaf expansion and declines thereafter. The length of the growing season before leaf senescence is, therefore, a limitation on crop yield (H. Thomas, WPBI, Aberystwyth). The efficiency of photosynthesis in relation to the balanced excitation of the photosystems is believed to be controlled by the phosphorylation of the light-harvesting chlorophyll *a/b* binding protein. Balanced excitation is necessary for efficient non-cyclic electron transport and also serves to poise cyclic electron flow by increasing excitation of photosystem I relative to photosystem II. Furthermore, the flexible distribution of available quanta between the two photosystems can be especially important under stress and allow a more efficient distribution of excitation in different cells of a leaf and different

*The ARC discussion meeting on 'What limits photosynthesis' was held on 15-16 April 1982 and was hosted by the ARC Research Group on Photosynthesis at the University of Sheffield.

New views of the immunoglobulin heavy-chain switch

from Kenneth B. Marcu and Max D. Cooper

DURING the differentiation of B lymphocytes, the immunoglobulin genes they express undergo a succession of rearrangements. The last of these determines the class of antibody expressed — whether it is IgM, IgG, IgA and so on. Just which of these classes or isotypes will predominate in an antibody response depends, among other things, on the nature of the antigen and in particular on the dependence of the B-cell response on T-cell help. Both the mechanism of the class switch and the nature of the T-cell influence were discussed at a meeting* in Sweden.

Antibody molecules (or immunoglobulins) consist of two identical heavy and light chains. In the mouse, these are encoded by linked families of V_H , D_H and J_H genes located 5' to a cluster of eight heavy-chain constant-region (C_H) genes (μ , δ , γ_3 , γ_1 , γ_2b , γ_2a , ϵ and α) on chromosome 12 and V_L and J_L genes upstream from the κ and λ light-chain genes on chromosomes 6 and 16 respectively.

One of the first steps in differentiation of the B cell is the assembly by chromosomal rearrangement of one each of the V_H , D_H and J_H genes and the transcription of the V-D-J set along with the C_μ gene (see the figure). A cell expressing such μ mRNA is known as a pre-B cell, and at this stage few of the μ chains reach the cell surface. Thus the pre-B cell is not influenced by antigen until the next stage in differentiation, when one set of V_L and J_L genes is productively rearranged and a complete IgM molecule is expressed on the surface of the cell, which has now differentiated into an immature B lymphocyte.

Subsequent stages in B-cell differentiation entail changes in the expression of the C_H genes, and, in particular, isotype switching in which the same V_H - D_H - J_H set is expressed sequentially with different constant-region genes. Intermediate stages in B-lymphocyte differentiation are marked by the expression of a variety of cell-surface proteins involved in regulating B-cell migration, growth and differentiation. In its terminally differentiated state, the B cell, now a plasma cell, shifts from surface expression to secretion of its antibody.

All immature B lymphocytes at first express membrane-bound IgM, and later co-express surface IgD, sharing the same V_H - D_H - J_H and light chains. Some

members of each B-cell clone make a further switch, from IgM (and IgD) to IgG, IgA or IgE — again with no change in the V-D-J genes. Heavy-chain isotype switching in B-cell clones was discovered more than a decade ago, when it was shown that treatment of immature B cells by infusions of anti- μ antibodies prevented subsequent development of plasma cells producing other immunoglobulin classes¹. It is now known that in plasmacytomas secreting isotypes other than μ , the constant-region genes 5' to the one that is expressed are deleted, the deletion occurring between switch (S) regions localized in the introns 5' to the expressed C_γ , C_ϵ or C_α gene²⁻⁶. But many fundamental questions remain unanswered. What is the normal sequence of C_H gene expression — can cells switch directly from IgM to, say, IgE or must they express each successive isotype between? What are the precise switch mechanisms and when does a cell become committed? How is the switch regulated? The answers to these questions are beginning to emerge and were discussed at Umeå.

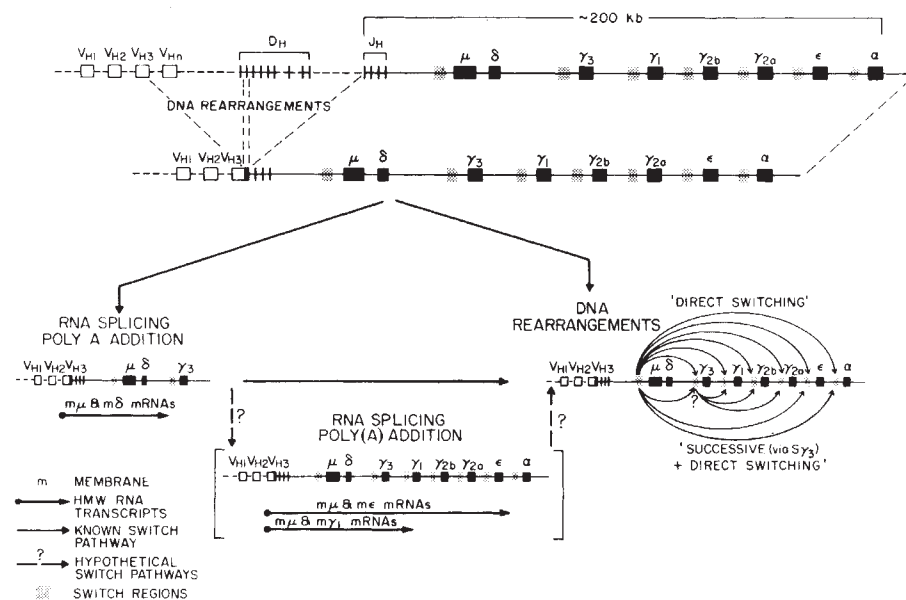
Mechanism of the switch

While it is generally accepted that deletion of the 5' constant-region genes accompanies differentiation into an immunoglobulin-secreting plasma cell, it is still not known whether this occurs by

intrachromosomal deletion² or by unequal exchange between sister chromatids⁷. In either case, there is now evidence that the C_H gene deletion may not occur until the switch from the surface to secreted immunoglobulin has taken place.

The existence of immature B lymphocytes exhibiting two cell-surface isotypes, for example IgM + IgG or IgM + IgA, has been known for some time. It was originally thought that this could be most simply explained by the stability of either the μ mRNA or the μ polypeptide, which would allow for the maintenance of cell-surface IgM for a B cell that has already genetically switched to either γ or α isotype. In the case of cells co-expressing μ and δ chains, however, it is now known that the two chains are produced by differential splicing of a single primary RNA transcript containing both the C_μ and the C_δ gene⁸⁻¹⁰. At the meeting it became clear that differential splicing of a long transcript may also underlie the expression of constant-region genes downstream of δ . T. Honjo (Osaka University) presented evidence that $\mu^+ \epsilon^+$ B lymphocytes, purified from the spleens of parasite-infected SJA/9 mice (a congenic strain without detectable circulating IgE antibody), have no detectable C_H gene rearrangements and/or deletions. To explain this unexpected finding, Honjo proposes²³ that a transcription unit starting 5' of V_H

MOUSE HEAVY CHAIN ISOTYPE SWITCH PATHWAYS



*A workshop on 'Immunoglobulin Heavy-Chain Class Switching' was organized by A. Coutinho and his colleagues at Umeå University, Sweden, on 16-19 February 1982.

Kenneth B. Marcu is at the Department of Biochemistry, State University of New York at Stony Brook, New York 11794, and Max D. Cooper is at the Cellular Immunobiology Unit, University of Alabama Medical Center, Birmingham, Alabama 35294.

In the course of B cell ontogeny, V_H , D_H and J_H DNA segments first become assembled to encode a complete heavy-chain variable region and later, the same assembly of V, D and J segments is associated sequentially with different C segments. The diagram shows the three possible ways in which this sequential switching may be achieved.

and terminating past ϵ is produced in $\mu^+ \epsilon^+$ lymphocytes. This approximately 180 kilobase multi- C_H transcript could yield μ and ϵ membrane mRNAs by differential RNA splicing coupled with specific poly(A) addition (see the figure). Similar studies on a population of $\mu^+ \gamma 1^+$ B cells from normal mouse spleen have been interpreted in the same way¹¹. However, cellular immunologists will want clear assurance that the immunoglobulin molecules on the surface of these cells are actually synthesized by them and not picked up from other cells, especially since the existence of a multi- C_H transcript in immature B lymphocytes may be very difficult to establish (none has yet been isolated for the μ - δ transcription unit). This hypothesis poses the intriguing problem of a mechanism for operating the switch first by RNA splicing and then by DNA rearrangement since γ^+ and α^+ B cells have been shown to give rise to plasma cells which produce the same heavy chain isotype.

Another hypothesis arises from observations on subclones of a pre-B cell line (18-81) transformed by Abelson virus. In these cells, switching from μ to $\gamma 2b$ (ref. 12) can apparently occur in the absence of μ -gene deletion, though not without some DNA rearrangement¹³. It is possible that at some point in the history of the 18-81 cell line, the δ , $\gamma 3$ and $\gamma 1$ genes were deleted and the $\gamma 2b$ gene was positioned immediately 3' to the μ gene where it could give rise to a μ - $\gamma 2b$ transcript by the same mechanism that produces the μ - δ transcript. F. Blattner (University of Wisconsin) suggested that this might be a transitional step in isotype switching, the constant-region gene to be expressed in each case replacing the δ gene 3' to μ . The next step would involve C_μ gene deletion before plasma cell differentiation.

Successive and direct switching

So far, there is evidence for both successive and direct isotype switching (see the figure). Both switching modes could operate on a purely stochastic basis⁵, or under the influence of regulatory elements such as isotype-specific recombinases¹⁴.

In addition to direct $\mu \rightarrow \gamma$ and $\mu \rightarrow \alpha$ switches, M. Wabl (University of Tübingen) reported evidence for a $\mu \rightarrow \epsilon$ switch in lipopolysaccharide (LPS)-induced plasma blasts. Successive isotype switches are also suggested from studies using the splenic focus assay and limiting dilution as methods for examining these cell clones^{15,16}. Using the limiting dilution of LPS-stimulated cells, A. Coutinho and co-workers (Umeå University) obtained results consistent with switches from μ to each of the other isotypes and for multiple patterns of successive 'downstream' switches. In the response to trinitrophenol-Ficoll, a T-independent antigen, the frequency of γ isotypes expressed follows the C_H gene order (that is, $\gamma 3 > \gamma 1 > \gamma 2b > \gamma 2a$), and results of splenic focus analysis suggested that $\gamma 3$

cells may be pivotal for switches to other γ isotypes¹⁶. K. Marcu (SUNY, Stony Brook) showed that $S_{\gamma 3}$, S_α and S_ϵ possess extensive homology with S_μ , while $S_{\gamma 1}$, $S_{\gamma 2b}$ and $S_{\gamma 2a}$ have very limited S_μ homology, and suggested that this could provide a genetic basis for intermediate $S_{\gamma 3}$ switches to other γ isotypes (see the figure). Marcu also presented evidence for a new consensus DNA sequence, YAGGTTG, preferentially located 5' of C_H switch sites and repeated in S regions^{17,18}, which, together with previously identified common S sequences (GAGCT and GGGGT)¹⁹, may play a part in switch-site recognition¹⁷⁻¹⁹.

Successive C_H gene switches have recently been explored at the molecular level in C_H switch-variant cell lines^{20,21}. A switch from $\gamma 2b$ to $\gamma 2a$ in MPC-11 myeloma variants results in S region-mediated $\gamma 2b$ gene deletion²⁰. F. Sablitzky (Genetics Institute, University of Cologne) presented some startling results from correlations of C_H gene rearrangements and deletions with isotype switching in clones of spontaneous class-switch variants derived from the X-63 hybridoma¹⁵. In X-63 hybridoma cells, 'backswitches' (for example $\gamma 2a \rightarrow \gamma 1$) occur at 100-fold higher frequency than forward switches (for example $\gamma 1 \rightarrow \gamma 2b \rightarrow \gamma 2a$), which arise at frequencies of 10^{-7} . The unprecedented finding of C_H backswitching could be most simply explained by either trans-homologous chromosome exchange or the presence of duplicated γ genes generated by unequal cross-over during sister chromatid exchange. Even though these findings may partly reflect peculiarities of the cell lines, they do clearly demonstrate that successive switching — for example, $\gamma 1 \rightarrow \gamma 2b \rightarrow \gamma 2a$ — can occur.

Role of T cells in isotype expression

It is well known that the nature of the antigen and responding T cells or their helper and suppressor factors can selec-

tively influence the isotypes expressed by plasma cell progeny of responding B cells. J. Davies (Washington University, St Louis) reviewed evidence indicating that T-independent polysaccharide antigens elicit primarily μ and $\gamma 3$ antibodies while the T-dependent protein antigens preferentially induce μ , $\gamma 1$ and $\gamma 2$ antibodies in the mouse. Coutinho showed that the same hapten can elicit different antibody isotypes when coupled to different carriers which determine the nature of the T-cell help. Y. Rosenberg (MRC, Mill Hill) showed that the isotype-specific influence of antigen-stimulated T cells may spill over to cause bystander B-cell clones to undergo plasma cell differentiation, the T-cell influence presumably being mediated by soluble factors.

Three groups of investigators represented at the Umeå meeting have analysed the immunoglobulin isotypes expressed by mouse B cells when cultured with established T-cell lines (A. Coutinho; and L. Forni, Basel) or soluble T-cell factors (P. Tucker, Dallas; & E. Severinsson, University of Stockholm). In each system, the polyclonal B-cell activator LPS by itself induced $\gamma 3$ more than $\gamma 1$ plasma cells. The addition of activated T cells, or soluble factors released from them favoured the reverse pattern, $\gamma 1$ over $\gamma 3$. L. Forni's data suggested that LPS induces both $\gamma 1$ and $\gamma 3$ B-cell blasts to proliferate, but only the $\gamma 3$ blasts to differentiate into mature plasma cells; the subsequent addition of activated T cells, by contrast, reduces the frequency of $\gamma 3$ cells and induces $\gamma 1$ plasma cells. P. Tucker presented data on the state of the $\gamma 3$ and $\gamma 1$ mRNA species before and after the addition of a T-cell product (B-cell differentiation factor²²) to LPS-stimulated B cells. LPS alone induced $\gamma 3$ mRNAs, but very little $\gamma 1$ mRNA. Following the addition of the T-cell factor, $\gamma 1$ mRNAs were increased in quantity and $\gamma 3$ mRNAs reduced. But all these results, while tantalizing, leave open the central question: do T cells instruct B cells to switch isotypes or do they select isotype-committed B cells for preferential induction of terminal differentiation?

Clearly, we still have important lessons to learn about the genetic basis for C_H isotype switching and its regulation. While solutions appear almost within reach, these genetic events remain elusive as they occur transiently and in minor subpopulations of cells. Clones of normal B cells undergoing isotype switches are unavailable. The C_H switching phenomenon may continue to challenge both cellular and molecular biologists for some time

Corrigendum

In "Sons and daughters" from T.H. Clutton-Brock (*Nature* 298, 11; 1982) the figures shown for *Papio cynocephalus* (Table 2) as "Per cent males" are actually totals; the percentage of males would be 35 (for high maternal rank) and 68 (for low maternal rank).

1. Kincade, P.W. *et al. Proc. natn. Acad. Sci. U.S.A.* 67, 1918 (1970).
2. Honjo, T. & Kataoka, T. *Proc. natn. Acad. Sci. U.S.A.* 75, 2140 (1978).
3. Cory, S. & Adams, J.M. *Cell* 19, 37 (1980).
4. Rabbitts *et al. Nature* 283, 351 (1980).
5. Coleclough, C., Cooper, D. & Perry, R.P. *Proc. natn. Acad. Sci. U.S.A.* 77, 1422 (1980).
6. Shimizu, A. *et al. Cell* 28, 499 (1982).
7. Obata *et al. Proc. natn. Acad. Sci. U.S.A.* 78, 2437 (1981).
8. Liu, C.P. *et al. Science* 209, 1348 (1980).
9. Maki, R. *et al. Cell* 24, 353 (1981).
10. Moore, K.W. *et al. Proc. natn. Acad. Sci. U.S.A.* 78, 1800 (1981).
11. Perlmutter, A.P. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
12. Burrows, P.D. *et al. Proc. natn. Acad. Sci. U.S.A.* 78, 564 (1981).
13. Alt, F.W. *et al. Nature* 296 325 (1982).
14. Davis, M.M., Kim, S.K. & Hood, L. *Science* 209, 1360 (1980).
15. Gearhart, P.J. *et al. Proc. natn. Acad. Sci. U.S.A.* 77, 5424 (1980).
16. Mongini, P., Metcalf, L. & Paul, W. *J. exp. Med.* 155, 884 (1982).
17. Lang, R.B., Stanton, L.W. & Marcu, K.B. *Nucleic Acids Res.* 10, 611 (1982).
18. Marcu, K.B. *et al. Nature* (in the press).
19. Nikaudo, T., Kakai, S. & Honjo, T. *Nature* 292, 845 (1981).
20. Eckhardt, L. *et al. Proc. natn. Acad. Sci. U.S.A.* 79, 3006 (1982).
21. Beyreuther *et al. ICN-UCLA Symp. molec. cell. Biol.* 20, 229 (1981).
22. Isakson, P.C. *et al. J. exp. Med.* 155, 734 (1982).
23. Yaoita, Y. *et al. Nature* 297, 697 (1982).

REVIEW ARTICLE

Structure and variation of human α_1 -antitrypsin

Robin W. Carrell, Jan-Olof Jeppsson*, Carl-Bertil Laurell*,
Stephen O. Brennan, Maurice C. Owen, Lloyd Vaughan & D. Ross Boswell

Molecular Pathology Laboratory, Pathology Department (University of Otago), Christchurch Hospital, Christchurch, New Zealand, and *Department of Clinical Chemistry, University of Lund, Malmö General Hospital, S-214 01 Malmö, Sweden

The sequence of α_1 -antitrypsin is in keeping with its role as a tissue scavenger of leukocyte elastase. Two abnormal variants commonly present in Europeans cause a deficiency that predisposes them to a progressive loss of lung elasticity. The nature of the reactive centre helps explain why cigarette smoking greatly accelerates the onset and severity of this degenerative process to give the disease emphysema.

AS its name implies, α_1 -antitrypsin (α_1 -AT) is the major component of the α_1 electrophoretic band of the human plasma proteins and acts as an inhibitor of proteolytic enzymes. It inhibits serine proteinases and hence is also known as α_1 -proteinase inhibitor. Interest in this protein stems largely from the observation that its genetic deficiency¹ is accompanied by the early onset of degenerative lung disease² and, in some instances, liver disease³. A puzzling feature of this genetic deficiency is its virtual confinement to Europeans, some 10% of whom are carriers of a pathological variant^{4,5}.

This association with disease has stimulated research into the structure and abnormalities of α_1 -AT that is revealing new aspects of the molecular pathology of genetic disease. One example^{6,7} is the way that genetic and environmental factors may interact to give disease. α_1 -AT is a protective factor and its genetic deficiency renders the individual susceptible to cumulative lung damage from bouts of inflammation induced by outside stress such as pollutants. The proteolytic enzymes released during this inflammation can overwhelm the limited amounts of inhibitor present in the deficient individual, a situation that is aggravated by the additional ability of some pollutants to produce a direct inactivation of α_1 -AT. The consequence is that whereas in a healthy environment α_1 -AT deficiency may be compatible with a full lifespan, in a polluted environment or with heavy cigarette smoking it can result in crippling lung disease and early death⁸.

This review discusses α_1 -AT, its variants and their molecular pathology, with particular reference to the primary structure of the protein presented here in its completed form.

Structure

The structure of human α_1 -AT shown in Fig. 1 is based on our amino acid sequence of the plasma protein and represents a consensus with that of the unpublished cDNA sequence of human α_1 -AT (S. L. C. Woo and E. W. Davie, personal communication). Details of the sequence and its determination are given in Fig. 1 legend.

The structural features of α_1 -AT are in keeping with its known functions. The molecule is relatively small and polar, allowing ready movement into the tissue fluids. It has a single polypeptide chain of 394 residues, with three carbohydrate side chains^{9,10}, giving an overall molecular weight of 51,000. The cDNA studies¹¹ confirm that α_1 -AT does not have a propeptide but does have a 24-residue hydrophobic pre- or signal peptide¹². The molecule has a single reactive site¹³, centred on methionine¹⁴ at position 358. For α_1 -AT to act as a proteinase inhibitor, this reactive site must be held in a fixed conformation: the conformation of an ideal substrate¹⁵. Although the tertiary structure of α_1 -AT is not known, physical measurements^{16,17} and calculations¹⁸ predict a highly ordered structure. The

strength of this ordered structure is indicated by the fact that it is not buttressed by the disulphide bridges that are characteristic of most inhibitors, including the homologous antithrombin-III¹⁹. α_1 -AT has only the one cysteine residue, located in a cavity on the surface of the molecule and available for SH-SS interchange reactions both *in vivo* and *in vitro*.²⁰

A characteristic feature of α_1 -AT is the multiple banding that it shows on acid electrophoresis (pH 4.9)²¹ or on electrofocusing²². This banding is the result of a microheterogeneity that is primarily due to variations in its carbohydrate structure²³. The three oligosaccharide groups of α_1 -AT are N-linked to asparagine residues at positions 46, 83 and 247.¹⁰ A physiologically controlled variation can occur in the structure of these side chains to give isoforms; physiological variants of α_1 -AT in which the usual biantennary oligosaccharides may be replaced in position 83, and to a lesser extent 46, by triantennary oligosaccharides^{9,23} (Fig. 2).

Reactive centre: Although it has been generally agreed that the reactive site of α_1 -AT is centred on a methionine residue¹⁴, its location in the sequence and the number of reactive sites has been controversial^{13,14,24-26}. The overall sequence shows that the reactive centre methionine is Met 358, 37 residues from the C-terminus of the molecule^{13,27,28}.

Proteinase inhibition is accompanied by the formation of a 1:1 complex between α_1 -AT and its target enzyme.²⁹ The reaction mimics that between the enzyme and its substrate, with the reactive centre of the inhibitor functioning as a bait for the active site of the proteinase. The mechanism of inhibition appears to involve a blockage of the reaction, probably at the stage of the tetrahedral adduct, and the evidence suggests that cleavage does not occur in physiological conditions²⁶. Conversion of the adduct to an acyl-enzyme may be induced by gross changes in pH or by treatment with strong denaturing agents; only in these non-physiological conditions can the existence of the C-terminal, 36 residue, leaving group be demonstrated²⁸.

Homology: An interesting feature of the sequence of α_1 -AT is its homology with those of antithrombin-III^{19,30} and ovalbumin³¹ (Fig. 1). It shares 29% common structure with antithrombin-III and 30% with ovalbumin, and ovalbumin and antithrombin-III share 32% common structure with each other. Together the three form a new superfamily³¹ that probably diverged from a common ancestral proteinase inhibitor some 500 million years ago. However, this suggestion of a common origin has been questioned as a consequence of gene sequence studies³² that indicate a complete difference in intron positions between α_1 -AT, with three introns, and ovalbumin, with seven introns. This has raised the possibility that the similarity of the two proteins may be due to convergent evolution. However, comparison of aromatic amino acids (Fig. 1) confirms that the structural similarity of α_1 -AT and ovalbumin extends

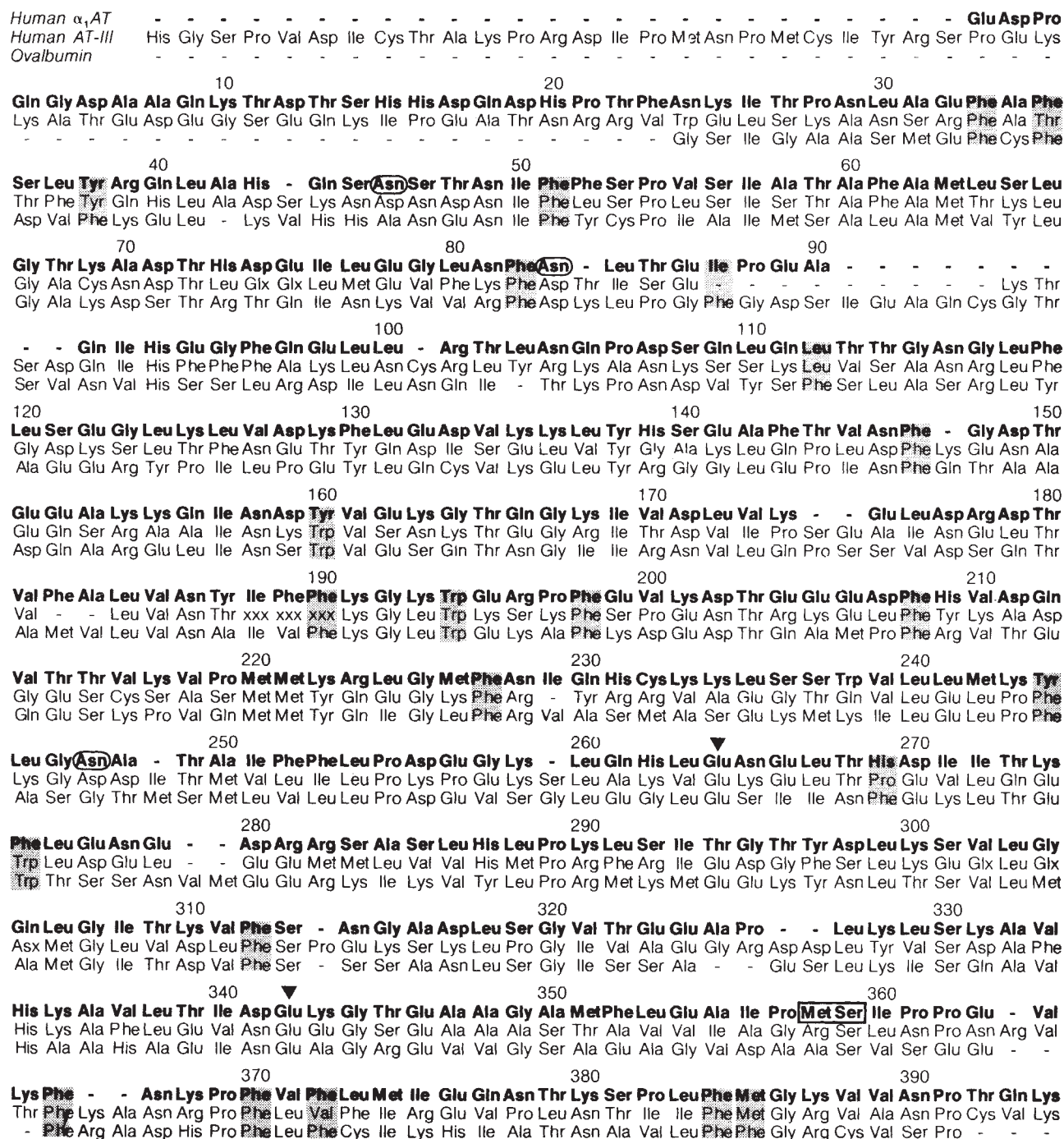


Fig. 1 The sequence of human α_1 -AT aligned with that of human antithrombin-III¹⁹ and chicken ovalbumin⁶⁸. Sequence numbers refer to α_1 -AT. The reactive centre Met-Ser 358-359 is boxed; the attachment points of the three oligosaccharide chains at Asn 46, Asn 83 and Asn 247 are encircled; ▼ shows the mutation sites of the S (264 Glu → Val) and Z (342 Glu → Lys) variants. Residues aligned with the bulky aromatic Phe and Trp of ovalbumin are shaded to indicate that structural homology is not confined to a single domain of the molecule; this confirms a relationship derived from divergent evolution. Homology of the reactive centre region 358-363 of α_1 -AT and antithrombin-III is illustrated in Fig. 3; ovalbumin retains some homology at this site, although it apparently has no inhibitory function. There has also been some conservation in α_1 -AT of the internal site (244-261) identified in ovalbumin as a hydrophobic signal sequence⁷¹, but α_1 -AT has an undoubted N-terminal signal sequence^{11,12}. There are several striking differences as well as the difference in intron numbers and positions. There are three disulphide bonds in antithrombin-III, one in ovalbumin and none in α_1 -AT; ovalbumin has a single oligosaccharide group, antithrombin-III has four and α_1 -AT has three, none of the sites being homologously related. α_1 -AT has a calculated molecular weight of 51,000 with a 13% carbohydrate content (isoform I). The sequence of human α_1 -AT is based on our previously published peptide sequences^{10,27,30} along with recent completion of overlap sequences. The sequence has been checked against that of the unpublished cDNA sequence for human α_1 -AT (S. L. C. Woo and E. W. Davie, personal communication). Residue Leu 176, and amide assignments 158 and 159, have been made on the basis of the cDNA structure. The only sequence difference is at 213 where we obtain an unequivocal Val but the cDNA shows an Ala; this may represent a polymorphism. Our sequence gives a consistent Asp-Gly at 116-117 but post-synthetic deamidation is known to occur at Asn-Gly sequences^{69,70} and residue 116 has therefore been assigned as an Asn, as found in the cDNA.

throughout the sequence. This would be highly improbable if the similarity were due to convergence.

The relationship of the two plasma proteins, α_1 -AT and antithrombin-III, is not surprising. Both are serine proteinase inhibitors and both have become specialized to varying degrees as inhibitors of specific enzymes. This is most marked with antithrombin-III, which is a specific inhibitor of the coagulation proteinases. By comparison, α_1 -AT is relatively unspecialized and will inhibit all serine proteinases including thrombin, its most effective target being leukocyte elastase^{33,34}. An examination of the differences between the two inhibitors provides a clue as to the process involved in their specialization. Calculation of the evolutionary distance between α_1 -AT and antithrombin-III gives a value of 160 accepted point mutations per 100 residues, which is similar to that between thrombin and elastase¹³. This similarity suggests that the specialization of the inhibitors has occurred in response to, and in concert with, the specialization of their target enzymes.

The alignment of the two sequences provides strong evidence for the positioning of their reactive centres, giving support²⁷ for a reactive centre in antithrombin-III at Arg 384. The similarity between these two proteins is an example of divergent evolution but, interestingly, the recognition of the sequence of the reactive centre of α_1 -AT was greatly aided by a set of convergent homologies¹⁴. Figure 3 shows the remarkable similarity between the reactive centres of both proteins and those of the unrelated plant proteinase inhibitors.

Other features of the homology of the proteins are indicated in Fig. 1. One point of interest is that the 32-residue N-terminal portion of α_1 -AT, which bears no resemblance to the other two proteins¹⁰, is homologous to another plasma proteinase inhibitor, α_1 -antichymotrypsin³⁵. This has a similar function to that of α_1 -AT as an inhibitor of a leukocyte enzyme, probably cathepsin G³³.

Function

There is good reason to believe that α_1 -AT is primarily a defence protein whose function is to protect the tissues against attack from released proteolytic enzymes. It is synthesized in the liver and secreted into the plasma where it has a half life of 6 days³⁶. As would be expected of a defence protein, it is an acute phase reactant and its normal plasma concentration³⁷ of 1.3 g l⁻¹ can quadruple in response to the general stimulus of inflammation or oestrogens. This rise in plasma concentration is due to an increased synthesis of α_1 -AT, particularly of its triantennary isoforms³⁸ (Fig. 2). The only other inhibitor present in blood in concentrations approaching that of α_1 -AT is the much larger α_2 -macroglobulin. Both inhibitors function as suicidal proteins that are taken up by the cells of the reticuloen-

	P ₁	P ₁ '	P ₂	P ₃	P ₄	P ₅
Human α_1 -AT (Elastase)	... Met	Ser	Ile	Pro	Pro	Glu ...
Human AT-III (Thrombin)	... Arg	Ser	Leu	Asn	Pro	Asn ...
Soybean (Trypsin)	... Lys	Ser	Asn	Pro	Pro	Gln ...
Garden Bean (Elastase)	... Ala	Ser	Ile	Pro	Pro	Gln ...
Lima Bean (Trypsin)	... Lys	Ser	Ile	Pro	Pro	Gln ...

Fig. 3 The reactive centres of α_1 -AT and antithrombin-III bear a striking similarity to those of the much smaller and unrelated plant proteinase inhibitors¹⁴. The P₁ residue reflects the specificity of the target enzyme. Elastase preferentially cleaves at small hydrophobic residues such as alanine (as in the garden bean inhibitor) and valine. Methionine is found in some other inhibitors of elastase¹⁵; it probably confers a wider specificity but also its lability may be of advantage in allowing the deactivation of the inhibitor.

dothelial system once they have formed complexes with their target proteinases³⁹.

The evidence that the prime target of α_1 -AT is leukocyte elastase is circumstantial but persuasive. The strongest evidence is the consequences of deficiency; the major effect is a progressive loss of elasticity in the lung, a loss that can be mimicked in model systems by the administration of elastase aerosols⁴⁰. Secondary evidence is provided by the measurement of the association rate constants of α_1 -AT with a range of proteinases³³. These give a decreasing order with: leukocyte elastase > chymotrypsin > cathepsin G > anionic trypsin > cationic trypsin > plasmin > thrombin. These association rate constants together with clinical and *in vivo* experimental observations indicate that α_1 -AT is a physiologically effective inhibitor of leukocyte elastase^{39,41} but a relatively ineffective inhibitor of trypsin, plasmin and thrombin.

An initially puzzling aspect of the role of α_1 -AT as a specific inhibitor of elastase is the occurrence of methionine at its reactive centre. Methionine in a Pro-Met sequence will provide a susceptible point for cleavage by elastase but valine⁴² or alanine would seem to be a better and more specific choice. Furthermore, methionine has the apparent disadvantage of being readily oxidized to its sulphoxide which is polar and does not provide a favourable proteolytic cleavage site for elastase^{43,44}. Consequently, α_1 -AT is readily inactivated by oxidizing agents such as the oxygen radicals released by phagocytes^{45,46}. However, a mechanism can be proposed in which this lability may provide a self-regulatory contribution to the process of inflammation.

Consider the formation of an abscess around a small foreign particle. An aggregation of leukocytes occurs to surround the particle and perform two separate functions—a discharge of bactericidal oxygen radicals and a release of granulocytic enzymes, including⁴⁷ elastase and cathepsin G. These enzymes facilitate the breakdown of adjacent connective tissue to give an area of liquefaction which isolates the foreign body in readiness for ejection. The spread of this liquefactive process by diffusion of the enzymes will be constrained by tissue proteinase inhibitors, predominantly α_1 -AT. This will limit tissue breakdown at the periphery of the inflammation, but allow it to occur at the focus of activity where α_1 -AT will itself be inhibited by released oxygen free radicals. In this way the lability of the methionine at the reactive centre of α_1 -AT could be advantageous, if it allowed the inhibition to be switched off at the focus of inflammation but continued to provide protection for outlying tissue.

However, the lability of α_1 -AT to oxidation will be distinctly disadvantageous in exposed situations such as occur in the fluids lining the respiratory tract. The elastic tissue of the lung is protected from surface attack primarily by two inhibitors: human bronchial inhibitor which predominates in the fluid lining the upper respiratory tract^{34,41} and α_1 -AT which predominates in the lower respiratory tract^{7,34,41}. Normally there is ample inhibitory capacity but the defences of the lower respiratory tract, because they depend on α_1 -AT, are vulnerable to combined oxidative and elastolytic attack, as is believed to occur as a consequence of heavy cigarette smoking⁴⁸.

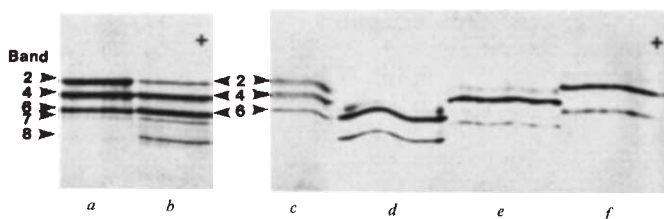


Fig. 2 Microheterogeneity of α_1 -AT has complicated the identification of its variants. This figure shows that the normal pattern (b, c) is formed by the superposition of three isoforms (d, e, f). The isoforms differ in the degree of branching of their oligosaccharide groups and consequently in the number of their negatively charged sialic acids. Isoform I (d) has all biantennary chains and isoforms II (e) and III (f) have one and two, respectively, of these replaced by triantennary chains. The pattern is complicated by a subsidiary heterogeneity seen in each isoform that is explicable by known minor post-secretory changes in primary structure. A shift to increased production of isoforms II and III during inflammation results in an increased proportion of bands 2 and 4 (a). (Reproduced with permission from ref. 38.)

Variation and disease

Genetic variation: There are several well established genetic polymorphisms of α_1 -AT. The protein is encoded by two independent alleles at a single locus, resulting in autosomal codominant inheritance. The variants are classified alphabetically by the *Pi* (proteinase inhibitor) nomenclature²¹, in terms of their electrophoretic mobility at pH 4.9. For example, *PiBB* is the homozygote for an anodal variant and *PiZZ* for a cathodal variant, with *PiMM* representing the homozygote for the normal *M* allele.

The variants of particular interest are the common *Z* and *S* mutants which are associated with plasma deficiency; the product of the *Z* allele gives concentrations equivalent to 15% of that of the normal *M* allele, whereas the product of the *S* allele gives concentrations of 60% of the normal allele⁴⁹. Mixed heterozygotes, for example *MZ*, *MS* and *SZ*, will give proportionately altered plasma levels. The genotypes with a deficiency sufficient to pose a definite risk to health are: *PiZZ* (15% of normal concentration), *PiSZ* (35% of normal) and combinations with the unusual null (–) gene, for example, *PiZ–* (8% of normal).

The structural abnormalities of the *Z* and *S* variants have been characterized^{50–52}. Both have single amino acid substitutions (Fig. 1) that have presumably arisen from single base mutations in the gene DNA. Otherwise the plasma forms of the two variants resemble the normal *M* protein; they have a similar half life in the circulation³⁷, normal inhibitory capacity and a similar carbohydrate composition. The *Z* protein, however, is something of an exception in that it is only partially secreted, and a considerable amount of the protein accumulates at the site of synthesis in the endoplasmic reticulum of the liver cells³.

A summary of the findings of various populations surveys⁵ shows that the *S* and *Z* variants are polymorphisms essentially confined to individuals of European descent. On average, some 3% of Europeans are carriers of the *Z* mutant (*PiMZ*) and some 7% are carriers of the *S* gene (*PiMS*). The frequency of the *Z* gene is somewhat higher in Northern Europeans and conversely that of the *S* gene is somewhat higher in Southern Europeans. About 1 in 1,000 Northern Europeans will have the *PiZZ* or *PiSZ* genotypes associated with severe deficiency of α_1 -AT⁵³. The relative confinement of these polymorphisms to the European is a puzzle. Both of the deficiency genes pose a real threat to health and there must be some balancing advantage to explain their survival at this frequency.

Deficiency and disease: Abnormal phenotypes of α_1 -AT have been associated with a number of diseases⁵⁴ including renal disease, arthritis and malignancies. However, the two strongest associations are with lung and liver disease; 60% of those with severe deficiency will die of lung disease and 14% of those with the *ZZ* deficiency will die of liver disease⁸. Two separate features need to be considered regarding the pathology of these diseases in α_1 -AT deficiency. These are the liver accumulation of the variant and the low plasma concentrations of inhibitor.

The accumulation of α_1 -AT in the hepatocytes occurs only with those variants whose defect is assumed to be one of secretion, the *Z* variant being by far the most common example⁵⁵. The *Z* protein that accumulates in the liver apparently has the same polypeptide structure⁵⁶ as the circulating *Z* protein, but its carbohydrate side chains are immature, with a high mannose content and no sialic acid^{56,57}. This is compatible with the deficiency of the *Z* protein being due to a partial block in processing of the polypeptide, with some of the material proceeding to completion and subsequent secretion in the plasma whilst the remainder is stopped in its passage through the secretory pathway at the stage of partial glycosylation⁵. The accumulation takes the form of intracellular amorphous aggregations of α_1 -AT that can be demonstrated microscopically as occasional deposits in the *PiMZ* liver and as massive deposits in the *PiZZ* liver (Fig. 4).

Liver disease is most apparent in the *PiZZ* newborn, where it occurs as a hepatitis with a wide variation in severity⁵⁸. Some

50% of *PiZZ* infants have minor changes in liver function, some 7% develop a severe obstructive jaundice, and a smaller proportion go on to develop a fatal juvenile cirrhosis. The reason for this variation in severity is not known; there are probably a number of contributory factors, such as hormonal influences, since boys are more affected than girls, and genetic influences, since a familial predisposition seems to be involved⁵⁸. Although the childhood cirrhosis is the most obvious and tragic form of the liver disease there is also an independent and substantial risk of developing cirrhosis in adult life. This increases with age to give cirrhosis levels near 20% in *PiZZ* individuals above the age⁸ of 50. The pathophysiology of this liver disease is not understood. It is almost certainly related to the accumulation of α_1 -AT in the hepatocyte rather than the associated plasma deficiency, because it only occurs in association with those variants that produce inclusions⁵ and not with those, such as the *S* variant, that give an isolated plasma deficiency.

However, there is no doubt that the lung disease is a direct consequence of low plasma levels of α_1 -AT. This conclusion is supported by a large body of evidence, both observational and experimental, that has been reviewed elsewhere^{6,7}. It all indicates that the loss of lung elasticity that occurs in severe α_1 -AT deficiency represents the summation of cumulative proteolytic damage. The end point of this loss of elasticity is the development of the irreversible lung disease, emphysema. This is a common disease and only some 1–2% of cases are due to a genetic deficiency of α_1 -AT; however, the recognition of this relationship has opened up an understanding of the pathogenesis of emphysema in general as being due to a 'protease-antiprotease' imbalance^{7,59}.

The most common cause of emphysema is heavy cigarette smoking. Exposure of the lungs to tobacco smoke causes a considerable increase in the population of a phagocytic cell peculiar to the lung, the pulmonary alveolar macrophage^{6,7}. The numbers of these in the lung increase up to 10-fold or more in proportion to the number of cigarettes smoked. The macrophages change in character and become activated to give increased proteolytic activity along with the secretion of a chemotactic factor that attracts leukocytes to the lungs⁷. This leads to a considerable change in the phagocytic population of the lung with a large increase in the elastolytic load. There is also an accompanying diminution of proteinase inhibitory capacity, due to an inactivation of inhibitor rather than a deficiency⁴⁴. Normal amounts of α_1 -AT are present in bronchial washings but its inhibitory capacity is almost halved⁴⁸. Some of this loss may be due to direct oxidation of the reactive methionine of α_1 -AT by cigarette smoke⁶⁰ but the main source of oxidation is almost certainly macrophage and leukocyte oxygen radicals^{46,61}. The conclusion is that a combination of

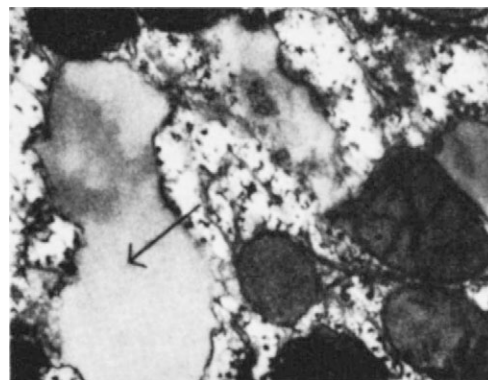


Fig. 4 Electron micrograph ($\times 26,500$) of a liver cell showing the accumulation of *Z* variant α_1 -AT in the rough endoplasmic reticulum. The position of the inclusions and their composition are compatible with a blockage occurring at the stage of final processing of the oligosaccharide side-chains. These inclusions are clearly visible by light microscopy and can be specifically identified³ with fluorescent antibody to α_1 -AT.

the increased production of elastase along with the decrease in inhibitor activity results in an attack on the elastic tissue of the lower respiratory tract to give the development of emphysema^{6,7,48}.

The heavy cigarette smoker with homozygous α_1 -AT deficiency is in double jeopardy and faces the likelihood of crippling emphysema before the age of 40 years⁶² with a 20-year decrease in survival time as compared with the non-smoking homozygote⁸. This situation illustrates the interaction that occurs between environmental and genetic factors. On the one hand there is the risk posed by macrophage stimulation due to smoking and air pollution; on the other, there is a whole range of genetic deficiencies of the proteinase inhibitor. For example, an intermediate deficiency such as *PiSZ* may be compatible with a fully active life for the non-smoking office worker⁵⁸, whereas it can result in premature emphysema in a heavily-smoking industrial worker. Fortunately, the common heterozygous genotypes, *PiMS* and *PiMZ*, pose no real threat to respiratory health^{63,64}.

Molecular pathology: There is now no doubt that the common *Z* and *S* deficiencies of α_1 -AT are due to single amino acid substitutions in its polypeptide chain. Independent analyses of plasma samples from several unrelated individuals have shown the change in structure in the *S* variant to be the replacement of glutamic acid by valine at position 264 and in the *Z* variant the replacement of glutamic acid by lysine at position 342^{18,50-52,65}.

In the absence of crystallographic data, the structural relationship of the two glutamic acids is unknown but an indication of their importance is their conservation in all the homologous proteins, including ovalbumin. The glutamic acid replaced in the *S* protein is in a predicted α -helical region⁵ and in this way it resembles the mutation of the coincidentally named haemoglobin S. Both involve the introduction of a valine on the hydrophilic aspect of a helical region, the effect of which will be to favour intermolecular aggregation. Support for this is seen in the decreased stability of the *S* variant of α_1 -AT⁶⁶. This decrease in stability suggests that deficiency may be due to increased breakdown, although no gross difference has been detected between the half life of the *S* and that of the normal M protein³⁶. The *S* protein does not accumulate in the liver and the problem is not apparently that of a failure of secretion, nor is it likely that a mutation of this type would affect translation rates. It is likely that the presence of the substituted valine results in incorrect folding of the nascent protein resulting in intracellular degradation and hence an apparent decrease in synthesis⁵.

The *Z* mutation presumably acts in a different way to give a partial blockage in the processing of the protein. The sequence of the protein excludes some of the previously suggested mechanisms, that is, that the mutation might directly interfere with pre- or propeptide cleavage or with an oligosaccharide attachment site^{3,37,67}. However, the sequence is of little help in suggesting alternative mechanisms. The mutation has occurred in a non-helical region but it must cause a perturbation in structure as the *Z* antitrypsin, like the *S*, has decreased *in vitro* heat stability⁶⁶. The failure in processing of the polypeptide could be a consequence of this perturbation in structure; alternatively, it has been proposed⁵ that it could be due to a misreading of the new Lys-Lys sequence 342-343 as a processing signal. No experimental evidence is available to support either proposal.

Conclusion

Knowledge of the structure of α_1 -AT with other available information provides at least a partial outline of the pathophysiology of the function and dysfunction of the protein. In particular, it supports the previously proposed concept^{6,7,59} that the emphysema associated with both α_1 -AT deficiency and cigarette smoking is due to a protease-antiprotease imbalance. This understanding of the pathogenesis of the lung disease provides the basis for the design of therapy either by anti-proteinase replacement or alternatively, by macrophage inhibition. One approach to replacement therapy is the development of low molecular weight analogues of the reactive centre of α_1 -AT⁴². A more radical approach will be the use of recombinant DNA techniques to give a substitution of the reactive centre methionine to decrease lability and increase specificity for elastase. But the most useful therapy that follows from knowledge of the pathogenesis of the lung disease is the concise and effective counsel that can be given to the *PiZZ* individual: Don't smoke!

The detection of homozygous deficiency at birth is technically and logistically a simple task.⁵³ The problem is to devise a means of contacting affected individuals at adolescence, at a time when they can elect not to smoke and thus have a reasonable chance of a full and active life.

This work was supported by the MRC of New Zealand and the Swedish MRC (project no. B81-13X-581-14C). We thank Dr S. L. C. Woo, Professor E. W. Davie and colleagues for access to the unpublished sequence of the cDNA of human α_1 -AT.

1. Laurell, C.-B. & Eriksson, S. *Scand. J. clin. Lab. Invest.* **15**, 132-140 (1963).

2. Eriksson, S. *Acta med. scand.* **175**, 197-205 (1964).

3. Sharp, H. L. *Hosp. Pract.* **6**, 83-96 (1971).

4. Fagerhol, M. K. *Postgrad. med. J.* **52** Suppl. 2, 73-83 (1976).

5. Carrell, R. W. & Owen, M. C. *Essays med. Biochem.* **4**, 83-119 (1979).

6. *Lancet* **i**, 743-745 (1980).

7. Gadek, J. E. *et al. Bull. Eur. Physiopath. Resp.* **16**, suppl., 27-40 (1980).

8. Larsson, C. *Acta med. scand.* **204**, 345-351 (1978).

9. Mega, T., Lujan, F. & Yoshida, A. *J. biol. Chem.* **255**, 4053-4056 (1980).

10. Carrell, R. W. *et al. FEBS Lett.* **135**, 301-303 (1981).

11. Kurachi, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6826-6830 (1981).

12. Carlson, J. & Stenflo, J. *FEBS Lett.* **130**, 297-300 (1981).

13. Boswell, D. R., Owen, M. C., Brennan, S. O., Carrell, R. W. & McLachlan, A. D. in *Proteins of the Biological Fluids: 27th Colloqu.* **28**, 129-132 (1980).

14. Johnson, D. & Travis, J. *J. biol. Chem.* **253**, 7142-7144 (1978).

15. Kraut, J. A. *Rev. Biochem.* **46**, 331-358 (1977).

16. Jirgensons, B. *Biochim. biophys. Acta* **493**, 352-358 (1977).

17. Saklatvala, J., Wood, G. C. & White, D. D. *Biochem. J.* **157**, 339-351 (1976).

18. Owen, M. C. thesis, Univ. Otago, Dunedin (1979).

19. Peterson, T. E., Dudek-Wojciechowska, G., Sottrup-Jensen, L. & Magnusson, S. in *The Physiological Inhibitors of Blood Coagulation and Fibrinolysis* (eds Collen, D. *et al.*) 43-54 (Elsevier, Amsterdam, 1979).

20. Laurell, C.-B. in *The Chemistry and Physiology of the Human Plasma Proteins* (ed. Bing, D. H.) 329-338 (Pergamon, New York, 1979).

21. Fagerhol, M. K. *Ser. Haemat.* **1**, Suppl. 1, 153-161 (1968).

22. Allen, R. C., Harley, R. H. & Talamo, R. C. *Am. J. clin. Path.* **62**, 732-739 (1974).

23. Vaughan, L. & Carrell, R. W. *Biochem. Int.* **2**, 461-467 (1981).

24. Martodam, R. R. & Liener, I. E. *Biochim. biophys. Acta* **667**, 328-340 (1981).

25. Kress, L. F., Kurecki, T., Chan, S. K. & Laskowski, S. M. *J. biol. Chem.* **254**, 5317-5320 (1979).

26. James, H. L. & Cohen, A. B. *J. clin. Invest.* **62**, 1344-1353 (1978).

27. Carrell, R. W., Boswell, D. R., Brennan, S. O. & Owen, M. C. *Biochem. biophys. Res. Commun.* **93**, 399-402 (1980).

28. Morii, M., Odani, S. & Ikenaka, T. *J. Biochem.* **86**, 915-921 (1979).

29. Cohen, A. B. *J. biol. Chem.* **248**, 7055-7059 (1973).

30. Carrell, R. W., Owen, M. C., Brennan, S. O. & Vaughan, L. *Biochem. biophys. Res. Commun.* **91**, 1032-1037 (1979).

31. Hunt, L. T. & Dayhoff, M. O. *Biochem. biophys. Res. Commun.* **95**, 864-871 (1980).

32. Leicht, M. *et al. Nature* **297**, 655-659 (1982).

33. Beatty, K., Bieth, J. & Travis, J. *J. biol. Chem.* **255**, 3931-3934 (1980).

34. Gauthier, F., Fryksmark, U., Ohlsson, L. & Bieth, J. G. *Biochim. biophys. Acta* **700**, 178-183 (1982).

35. Travis, J., Garner, D. & Bowen, J. *Biochemistry* **17**, 5647-5651 (1978).

36. Jeppsson, J.-O., Laurell, C.-B., Nosslin, B. & Cox, D. W. *Clin. Sci. molec. Med.* **5**, 103-107 (1978).

37. Jeppsson, J.-O., Laurell, C.-B. & Fagerhol, M. K. *Eur. J. Biochem.* **83**, 143-153 (1978).

38. Vaughan, L., Lorier, M. A. & Carrell, R. W. *Biochim. biophys. Acta* **701**, 339-345 (1982).

39. Balldin, G., Laurell, C.-B. & Ohlsson, K. *Hoppe-Seyler's Z. physiol. Chem.* **359**, 699-708 (1978).

40. Senior, R. *et al. Am. Rev. resp. Dis.* **116**, 469-475 (1977).

41. Gadek, J. E., Fells, G. A., Zimmerman, R. L., Rennard, S. I. & Crystal, R. G. *J. clin. Invest.* **68**, 889-898 (1981).

42. Nakajima, K., Powers, J. C. Ashe, B. M. & Zimmerman, M. *J. biol. Chem.* **254**, 4027-4032 (1979).

43. Schechter, Y., Burstein, Y. & Gertler, A. *Biochemistry* **16**, 992-997 (1977).

44. Beatty, K., Robertie, P., Senior, R. & Travis, J. *J. Lab. clin. Med.* (in the press).

45. Carp, H. & Janoff, A. *Am. Rev. resp. Dis.* **118**, 617-621 (1978).

46. Matheson, N. R., Wong, P. S. & Travis, J. *Biochem. biophys. Res. Commun.* **88**, 402-409 (1979).

47. Baglioni, M. in *The Cell Biology of Inflammation* (ed. Weissman, G.) 161-187 (Elsevier, Amsterdam, 1980).

48. Gadek, J. E., Fells, G. A. & Crystal, R. G. *Science* **206**, 1315-1316 (1979).

49. Laurell, C.-B. & Jeppsson, J.-O. in *The Plasma Proteins. Structure, Function and Genetic Control* Vol. 1, 2nd edn (ed. Putnam, F. W.) 229-264 (Academic, New York, 1975).

50. Owen, M. C. & Carrell, R. W. *Br. med. J.* **i**, 130-131 (1976).

51. Jeppsson, J.-O. *FEBS Lett.* **65**, 195-197 (1976).

52. Yoshida, A., Lieberman, J., Gaidulis, L. & Ewing, C. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1324-1328 (1976).

53. Laurell, C.-B. & Sveger, T. *Am. J. hum. Genet.* **27**, 213–217 (1975).
 54. Fagerhol, M. K. & Cox, D. W. *Adv. hum. Genet.* **11**, 1–62 (1981).
 55. Blenkinsopp, W. K. & Haffenden, G. P. *J. clin. Path.* **30**, 132–137 (1977).
 56. Jeppsson, J.-O., Larsson, C. & Eriksson, S. *New Engl. J. Med.* **293**, 576–579 (1975).
 57. Hercz, A., Katona, E., Cutz, E., Wilson, J. R. & Barton, M. *Science* **201**, 1229–1232 (1978).
 58. Sveger, T. *New Engl. J. Med.* **294**, 1316–1321 (1976).
 59. Lieberman, J. *Chest* **70**, 62–67 (1976).
 60. Janoff, A., Carp, H., Lee, D. K. & Drew, R. T. *Science* **206**, 1313–1316 (1979).
 61. Carp, H. & Janoff, A. *J. clin. Invest.* **63**, 793–797 (1979).
 62. Janus, E. D. & Carrell, R. W. *N. Z. med. J.* **81**, 461–467 (1975).
 63. Larsson, C., Eriksson, S. & Dirksen, H. *Br. med. J.* **ii**, 922–925 (1977).
 64. de Hamel, F. A. & Carrell, R. W. *N. Z. med. J.* **94**, 407–410 (1981).
 65. Yoshida, A., Ewing, C., Wessels, M., Lieberman, J. & Gaidulis, L. *Am. J. hum. Genet.* **29**, 233–239 (1977).
 66. Lieberman, J. *Chest* **64**, 579–584 (1973).
 67. Bell, O. F. & Carrell, R. W. *Nature* **243**, 410–411 (1973).
 68. McReynolds, L. *et al. Nature* **273**, 723–728 (1978).
 69. Moo-Penn, W. F. *et al. J. biol. Chem.* **251**, 7557–7562 (1976).
 70. Perutz, M. F., Fogg, J. H. & Fox, J. A. *J. molec. Biol.* **138**, 669–670 (1980).
 71. Lingappa, V. R., Lingappa, J. R. & Blobel, G. *Nature* **281**, 117–121 (1979).

ARTICLES

Internal rotation and gravitational quadrupole moment of the Sun

D. O. Gough

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

The internal angular velocity of the Sun is estimated from observations of apparent rotational splitting of low-order low-degree global oscillations detected in fluctuations in the limb-darkening function. A sidereal rotation of $3 \mu\text{Hz}$ is inferred for the deep interior; this is more than six times greater than the equatorial rotation frequency of the photosphere. The inferred angular velocity distribution does not vary strongly with latitude, and yields a gravitational quadrupole moment $J_2 \approx 3.6 \times 10^{-6}$. When combined with the results of planetary radar observations to determine $(2 - \beta + 2\gamma)/3$, a combination of Eddington–Robertson post-newtonian parameters which in general relativity is unity, a value of about $0.994 (\pm 0.012)$ is thus obtained.

ONE of the most important tests of theories of gravitation concerns the rates of precession of planetary orbits. In the fully conservative parameterized post-newtonian (PPN) formalism, for example, the predicted advance $\Delta\phi_0$ per orbital period of a planetary orbit with semi-major axis a and eccentricity e , after correcting for perturbations due to the other planets, is $6\pi GM\lambda_p/[a(1-e^2)c^2]$, where

$$\lambda_p = \frac{1}{3}(2 - \beta + 2\gamma) + \frac{R^2 c^2}{2GMa(1-e^2)} J_2 \quad (1)$$

Here M and R are the mass and radius of the Sun, G is the gravitational constant, c is the speed of light, and β and γ are Eddington–Robertson parameters of the PPN formalism^{1,2}. In general relativity, for example, $\beta = 1$ and $\gamma = 1$, and hence $\frac{1}{3}(2 - \beta + 2\gamma) = 1$. The parameter J_2 measures the quadrupole moment of the Sun's gravitation field, normally presumed to arise from centrifugal distortion. It is defined such that the solar component of the gravitational potential outside the Sun, referred to spherical polar coordinates (r, θ, ϕ) with respect to the Sun's rotation axis, is

$$\Phi = \frac{GM}{r} \left[1 - J_2 \left(\frac{R}{r} \right)^2 P_2(\cos \theta) \right] \quad (2)$$

where P_2 is a Legendre function of second degree. Equation (1) was derived assuming the Sun's rotation axis to be perpendicular to the plane of the planetary orbit. The correction that should be made to account for the small tilt that exists is insignificant.

Since the contributions to $\Delta\phi_0$ from the quadrupole component of Φ and from the deviations from newtonian gravitation depend differently on $a(1-e^2)$, one should in principle be able to determine $\frac{1}{3}(2 - \beta + 2\gamma)$ and J_2 separately by combining observations of different planets. However, it has not been possible to do this, because the planetary precession rates are not yet known with adequate precision³. An independent measurement of J_2 is therefore required.

Attempts have been made to infer J_2 from the shape of the visible solar disk. The method relies on equating the oblateness E of the disk with that of the gravitational equipotentials, after accounting for the observed rotation of the photosphere and brightness inhomogeneities such as are produced by faculae.

This yields $J_2 = \frac{2}{3}E$. It is interesting that modern observations have yielded quite different values: Dicke and Goldenberg⁴ obtained

$$\frac{2}{3}E = (24.7 \pm 2.3) \times 10^{-6} \quad (3)$$

whereas Hill and Stebbins⁵ obtained

$$\frac{2}{3}E = (1.0 \pm 4.3) \times 10^{-6} \quad (4)$$

The discrepancy has not been explained. Note, however, that the two measurements utilized different properties of the intensity distribution near the limb of the Sun, and that alone might prove to be the cause. Evidently one cannot infer J_2 until the discrepancy has been understood.

For the orbit of Mercury, equation (1) yields

$$\lambda_p = \frac{1}{3}(2 - \beta + 2\gamma) + 2.96 \times 10^3 J_2 \quad (5)$$

Shapiro *et al.*⁶, using radar measurements of the distance to Mercury, found

$$\lambda_p = 1.003 \pm 0.005 \quad (6)$$

A similar analysis by Anderson *et al.*³ gave

$$\lambda_p = 1.007 \pm 0.005 \quad (7)$$

Shapiro *et al.* took J_2 to be equal to the value one calculates assuming the Sun to be rotating uniformly, which is 2×10^{-7} if the surface equatorial angular velocity is used, and thus arrived at a value for $\frac{1}{3}(2 - \beta + 2\gamma)$. It is evident from the uncertainty suggested by the difference between equations (3) and (4) that the passage from λ_p to the PPN parameters is premature.

The limitations imposed by the uncertainty in J_2 on the accuracy to which the precession data test theories of gravity has led to other proposals for probing the solar gravitational field. One method is to take advantage of the high eccentricity of Mercury's orbit to search for different periodic orbital variations. This would require tracking a spacecraft in orbit around Mercury. According to Anderson *et al.*⁷ a precision of a few parts in 10^7 should be possible. Another suggestion is to track a spacecraft that passes by the Sun within a few solar radii⁸. Similar precision might be possible, though it could be affected by perturbations in Φ associated with long-period solar oscillations⁹, especially if J_2 were as small as 2×10^{-7} . Preliminary calculations^{7,10} suggest that J_2 could be measured to an accuracy

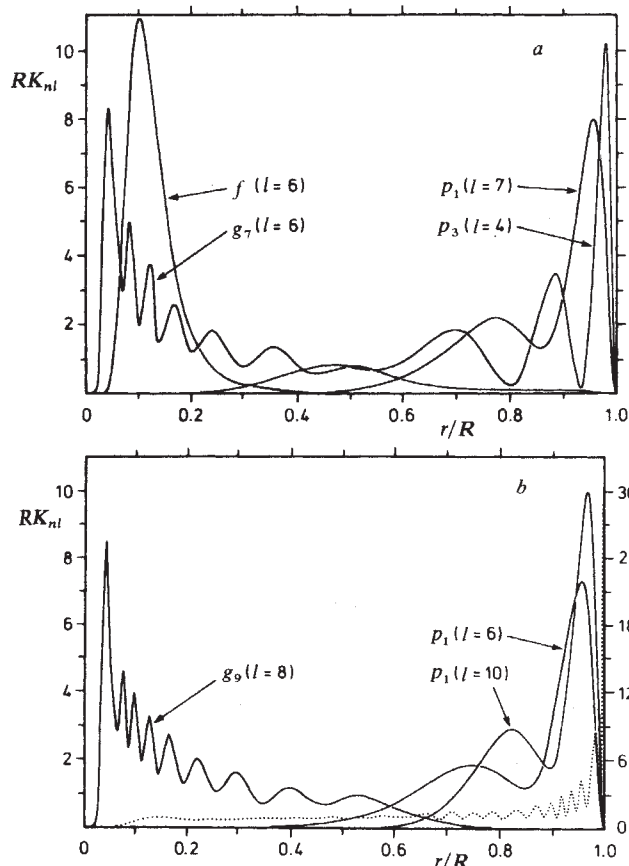


Fig. 1 Rotational splitting kernels RK_{nl} for the seven modes listed in Table 1. The dotted curve (to which the right-hand ordinate scale refers) is an average of the six kernels with $l=1$ and $l=2$ and frequencies closest to 2.5, 3.3 and 4.1 μHz . This is representative of the 5-min modes observed by Claverie *et al.*²⁵.

of a part in 10^8 if the spacecraft were in a bound orbit around the Sun.

The discovery of global solar oscillation by Hill and his collaborators¹¹, Severny *et al.*¹² and Brookes *et al.*¹³ raised the possibility that the deep interior of the Sun could be measured more directly than had previously been possible¹⁴. In particular, rotational splitting of otherwise degenerate nonaxisymmetrical modes provides averages of the angular velocity Ω . Different modes weight Ω differently, and with a sufficient variety one can go some way towards determining the variation of Ω with position. I report here an attempt to estimate Ω from recent observations. This yields a preliminary rough estimate of the Sun's angular momentum, the rotational kinetic energy and J_2 .

Properties of rotational splitting

Slow rotation causes nonaxisymmetrical standing wave patterns to precess about the rotation axis of the Sun. This is manifest

as a splitting of the degeneracy in frequency with respect to the angular order m of the modes^{15,16} (that is to say, the order m of the spherical harmonic that factors from the eigenfunctions). The splitting is an average of the angular velocity, weighted by a kernel that depends on the eigenfunction of oscillation of the nonrotating star. That kernel, which can be computed directly from the variational principle of Lynden-Bell and Ostriker¹⁷, is discussed in refs 18–20. Here I assume that the solar angular velocity Ω can be approximated by $\Omega\mathbf{k}$, where $\Omega(r, \theta) = \Omega_0(r) - \Omega_1(r) \cos^2 \theta$ and $\mathbf{k}$ is a constant unit vector. In that case the cyclic frequency $\nu \equiv \nu_{nlm}$ of a mode of order n and degree l , with respect to an inertial frame of reference, is well approximated by $\nu_{nl} + \nu'_{nlm}$, where ν_{nl} is the corresponding frequency when $\Omega = 0$, and

$$\nu'_{nlm} = \frac{m}{2\pi A} \int_0^R \rho \{ [(\xi - \eta)^2 + [l(l+1) - 2]\eta^2] \Omega_0(r) - [Q_{lm} \{ (\xi - \eta)^2 + [l(l+1) - 7]\eta^2 + \eta^2 \} \Omega_1(r)] r^2 dr \} \quad (8)$$

$$\equiv \frac{m}{2\pi} \left\{ \beta_{nl} \int_0^R K_{nl}(r) \Omega_0(r) dr - \beta_{nlm} \int_0^R K_{nlm}(r) \Omega_1(r) dr \right\} \quad (9)$$

$$A = \int_0^R \rho [\xi^2 + l(l+1)\eta^2] r^2 dr \quad (10)$$

$$Q_{lm} \equiv \frac{1}{2}(2l+1) \frac{(l-|m|)!}{(l+|m|)!} \int_{-1}^1 \mu^2 [P_l^m(\mu)]^2 d\mu \quad (11)$$

ξ and η , which depend on n and l but not on m , are measures of the radial and horizontal components of the adiabatic displacement eigenfunction $\delta\mathbf{r}$, defined in spherical polar coordinates by

$$\delta\mathbf{r}(r, t) = \text{Re} \left\{ \left[\xi(r) P_l^m(\cos \theta), \eta(r) \frac{d}{d\theta} P_l^m(\cos \theta) \right] \exp(im\phi - 2\pi i\nu t) \right\} \quad (12)$$

P_l^m is the associated Legendre function of the first kind, ρ is the equilibrium density of the Sun (which may be calculated ignoring the centrifugal distortion) and t is the time. The coefficients Q_{lm} can be expressed in terms of Gaunt integrals, which can be computed from the tables of Scott²¹. The factors β_{nl} and β_{nlm} are defined such that the kernels K_{nl} and K_{nlm} are unimodular (that is, their integrals are unity). Thus, if Ω_0 and Ω_1 were constant, the rotational perturbation ν'_{nlm} to the eigenfrequency would be $(m/2\pi)(\beta_{nl}\Omega_0 - \beta_{nlm}\Omega_1)$.

If the latitudinal variation of Ω in the solar interior is no greater than the variation observed in the photosphere, then for p modes and low-order g modes of low degree, the second integral in equation (9) is considerably smaller than the first. Thus, since β_{nl} and K_{nl} are independent of m , the frequencies corresponding to the $2l+1$ possible values of m are approximately uniformly spaced.

Observations of apparent rotational splitting

Oscillations of the radial variation of intensity near the solar limb have been recorded by Bos and Hill²². Measurements were

Table 1 Rotationally split multiplets and the modes with which they have been identified

Multiplet	$l_{\min}$	ν_c (μHz)	$\Delta \pm \sigma$ (μHz)	Mode	ν_{nl} (μHz)	$\beta_{nl}^{-1}(\Delta + \omega_e/2\pi)$ (μHz)
1	6	256.54	2.875 ± 0.008	$g_7(l=6)$	258.8	2.945
2	8	260.38	2.902 ± 0.006	$g_9(l=8)$	260.9	2.948
3	6	427.97	2.888 ± 0.008	$f(l=6)$	425.6	2.958
4	6	471.10	0.96 ± 0.02	$p_1(l=6)$	495.4	1.07
5	7	498.12	0.85 ± 0.02	$p_1(l=7)$	518.7	0.94
6	10	592.65	0.74 ± 0.01	$p_1(l=10)$	584.6	0.81
7	4	773.94	1.29 ± 0.01	$p_3(l=4)$	759.2	1.39

The theoretical eigenfrequencies ν_{nl} are the frequencies ν_λ extrapolated from the frequencies ν_A, ν_B of models A and B of ref. 30 with $\lambda = 1.26$. The corresponding frequencies of $f(l=32)$ and $p_2(l=10)$ are 588.5 and 780.1 μHz . Aside from possible errors in identification, the frequencies ν_c of the central components of the observed multiplets have standard error 0.07 μHz . The r.m.s. difference between ν_{nl} and ν_c is 13 μHz . Δ is the observed mean splitting, and σ is, according to Hill²⁴, a conservative estimate of the variation of the splitting with m . The last entry is numerically equal to the mean value of Ω_0 , weighted by the appropriate kernel K_{nl} illustrated in Fig. 1.

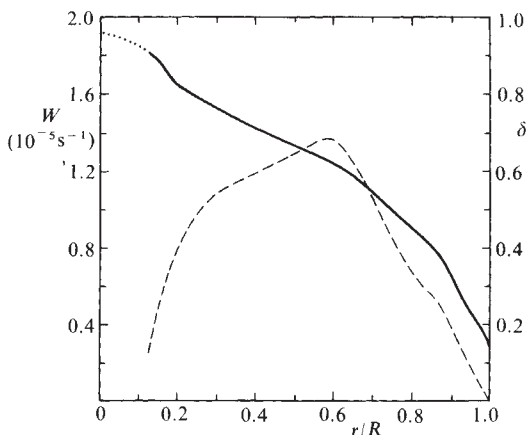


Fig. 2 Averages $W(r)$ of Ω_0 computed by the Backus-Gilbert optimal averaging procedure. Averaging kernels D with $\bar{r} < 0.12R$ cannot be constructed from the data kernels in Fig. 1, so W has been extrapolated to the origin with a cubic curve (dotted) that approximates W in the range $0.12 < r/R < 0.2$. Since $\mathcal{J}$ and $\mathcal{J}$ are both small near the origin (see Fig. 4), H , T , and J_2 are insensitive to the extrapolation procedure. The dashed line is the half-width δ of the averaging kernel D .

made simultaneously at three pairs of diametrically opposed locations, distributed symmetrically about the equator, and from them six functionals $\psi_i(t)$ of the limb-darkening function were constructed. For this discussion it does not matter how the functionals ψ_i are defined. It is necessary only to accept that their variation is produced by the ensemble of normal modes of oscillation of the Sun, whose frequencies are thus indicated by peaks in the power spectra of ψ_i . The solar origin of the signal is argued in refs 22, 23.

Observations of mean duration 9.3 h have been made on 18 days during a 41-day period²². Fourier transforms of $\psi_i(t)$ were computed, and from linear combinations of them power spectra of components of the oscillations satisfying certain symmetry relations could be isolated. For example, the power spectrum of $\sum_{i=1}^6 \psi_i$ contains contributions from only those modes with even l and even m . Part of that spectrum is shown in ref. 22. Aliasing aside, the frequency resolution is $1/(41 \text{ day}) = 0.28 \mu\text{Hz}$.

In the hope that the influence of the latitudinal variation of Ω on the splitting would be small, sequences of uniformly spaced peaks in the spectra were sought. Some care was taken to account for the side-bands imposed by the window function, but a systematic attempt to remove them was not made. To be acceptable, the spacing between the centres of the peaks was required to be uniform to $0.07 \mu\text{Hz}$. It is prudent to apply so stringent a condition, at least in a preliminary analysis, to guard against misinterpreting accidental sequences of peaks in the densely populated spectra. Seven sequences have been found. Their median frequencies ν_c , the mean frequency separations Δ and lower bounds $l_{\min}$ to l , inferred from the number of peaks in the sequences, were announced recently by Hill²⁴ and are reproduced in Table 1. For this investigation, I shall assume, with Hill, that these are sequences of rotationally split eigenfrequencies, and ask what that implies.

Before this analysis of the power spectrum, evidence of a mean rotational splitting of $0.75 \mu\text{Hz}$ synodic in the low-degree 5-min oscillations was discovered by Claverie *et al.*²⁵. The data are difficult to interpret because $2l+1$ components appear in the spectrum, whereas, owing to the symmetry of observing technique, only $l+1$ of the modes of degree l should be visible. Isaak²⁶ has suggested that the extra components are accounted for by an intense rotating magnetic field in the core of the Sun, as has been postulated by Dicke^{27,28} to explain the 12.2-day variation in the Princeton oblateness data. However, it is not clear how that explains the 5-min oscillation data²⁹. It is conceivable that a weak field could couple nearly resonant modes with different spatial symmetries, but in that case there is the danger that the coupling modifies the frequencies sufficiently to belie

the apparent splitting that arises directly from rotation. Therefore a detailed account of the analysis that includes the 5-min modes will not be presented.

Identification of the modes

Tentative identifications were made by selecting those modes of the standard (nonrotating) solar model A of Christensen-Dalsgaard *et al.*³⁰ with l approximately equal to but not less than $l_{\min}$ whose frequencies are closest to the median frequencies of the groups of modes observed. It was borne in mind that if $l_{\min}$ underestimates l , as a result of failure to detect or recognize modes with $|m| \approx l$, the median frequency may differ from the frequency ν_{nl0} by an integral multiple of the frequency separation Δ . The identifications are listed in Table 1. In all cases $l = l_{\min}$.

It is important to realize that there is considerable uncertainty in the identification. This is true particularly of the multiplets 1–3 and multiplets 6 and 7. In the middle range there is less scope for error because the eigenfrequencies are more sparsely distributed.

Splitting kernels

Kernels $K_{nl}(r)$ are illustrated in Fig. 1 for the modes listed in Table 1. It is a property of all these modes that for given n and l the kernels K_{nlm} are nearly the same as K_{nl} . Therefore the kernels K_{nlm} are not displayed. This property also characterizes those other modes that might have been identified with the observed oscillations, and arises because the motion is nearly vertical wherever the amplitude of the mode is large.

It is immediately evident from Fig. 1 that the low-frequency modes sample mainly the rotation of the core, and the high-frequency modes the rotation near the surface. The greater values of Δ associated with the lower frequencies indicates immediately that the Sun's core is rotating more rapidly than the photosphere.

Latitudinal variation of Ω

Because, for the modes under consideration, $l(l+1)\eta^2$ is rather smaller than ξ^2 , and $K_{nlm} \approx K_{nl}$, the splitting formulae (8), (9) may be approximated by

$$\nu'_{nlm} \approx \frac{m}{2\pi} \beta_{nl} (\bar{\Omega}_0 - Q_{lm} \bar{\Omega}_1) \quad (13)$$

Table 2 Properties of estimates of Ω

Estimate	H	T	J_2	σ_Δ
1	5.0	25.7	3.6×10^{-6}	0.07
2	4.8	23.8	3.8×10^{-6}	0.01
3	5.3	31.5	4.3×10^{-6}	0
4	5.9	37.4	5.4×10^{-6}	0.04
5	5.4	31.3	4.1×10^{-6}	0.02
6	4.4	27.9	4.5×10^{-6}	0
7	1.0	3.8	1.4×10^{-6}	0

Estimate 1 is the result of the Backus-Gilbert inversion shown in Fig. 1. Estimates 2–5 are the curves 1–4 in Fig. 2. Estimate 6 is the seventh-degree polynomial in r with $\Omega(R) = \Omega_S$ that satisfies the splitting constraints (14) exactly. Estimate 7 is the function Ω_0 that minimizes J_2 and satisfies $\Omega_0(R) = \Omega_S$ and the splitting constraints (14). For any value J above the minimum J_2 there are infinitely many functions $\Omega(r)$ satisfying $J_2 = J$, the outer boundary condition and the splitting constraints (14). The entries σ_Δ are r.m.s. differences between Δ and the frequency splitting predicted by Ω_0 , measured in units of the observed splitting Δ . The angular momentum H and the rotational kinetic energy T are measured in units of $I\Omega_S^2$ and $\frac{1}{2}I\Omega_S^2$ respectively, where $I = 0.068MR^2$ is the moment of inertia of the extrapolated solar model about an axis through the centre. If the rotational splitting of low-degree 5-min modes claimed by Claverie *et al.*²⁵ is added to the data, the values of $10^6 J_2$ for the seven estimates of Ω listed above are 2.3, 2.1, 3.4, 2.9, 2.8, 14.4 and 1.4 respectively. In this case estimate 6 is the eighth-degree polynomial that satisfies the constraints. If the last two modes in Table 1 are replaced by $f(l=32)$ and $p_2(l=10)$, whose splitting kernels are somewhat more concentrated near the surface, the corresponding values of $10^6 J_2$ are 3.6, 2.1, 5.1, 5.9, 4.2, 438.4 and 1.2 when the 5-min datum is ignored, and 2.4, 1.7, 4.9, 2.9, 2.9, 790.3 and 1.2 when the 5-min datum is included.

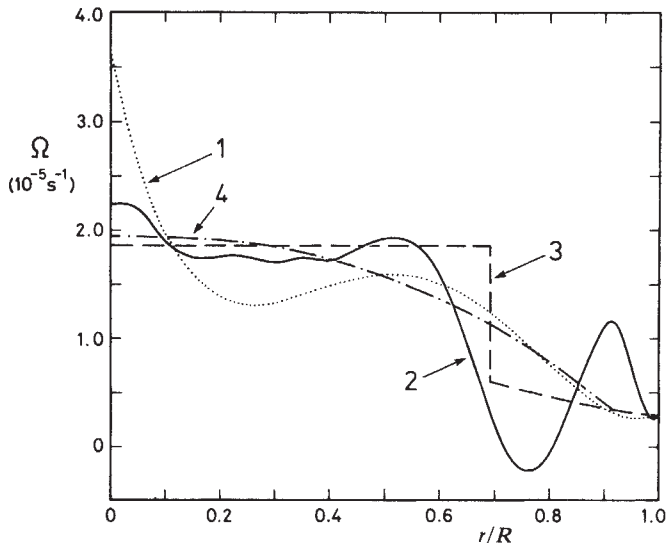


Fig. 3 Some possible rotation curves Ω_0 that are compatible with the data. Curve 1 is a fourth-degree polynomial and curve 2 minimizes the mean square gradient of Ω_0 . Curves 3 and 4 vary inversely with r^2 for $r > r_c$; for $r < r_c$ curve 3 is constant whereas curve 4 is proportional to $1 - \alpha(r/R)^b$. In each case the free parameters were chosen to fit to the constraints (14) by least squares, subject to curve 4 being continuous. Curve 2 satisfies (14) exactly. The data in Table 1 cannot distinguish between these curves.

where the overbar denotes the mean with weight K_{nl} . Thus from any observed nonuniformity in the splitting an estimate of $\bar{\Omega}_1/\bar{\Omega}_0$ may be obtained.

Nonuniformity in the splitting has several origins. Aside from observational error and differential rotation, the most likely source is a magnetic field. This can be distinguished from rotational splitting by a difference of symmetry. If, for example, ν_{nlm}^+ is the magnetic perturbation to frequency due to an axisymmetrical field, then $\nu_{nl(-m)}^+ = \nu_{nlm}^+$, whereas $\nu_{nl(-m)}^- = -\nu_{nlm}^-$. This result holds even when the axis of symmetry does not coincide with the rotation axis, provided ν_{nlm}^+ is less than the width of the lines in the power spectrum. With sufficiently accurate data the two contributions could be separated.

It is a consequence of the different symmetries that ν_{nlm}^+ and ν_{nlm}^- cannot cancel for all m . Therefore the deviations σ from uniform splitting listed in Table 1 can be used to estimate an upper bound to $\bar{\Omega}_1$. The standard deviation of the rotational contribution to the nonuniformity in the splitting can be computed easily from the approximation (13). If all detectable modes are present, the result, expressed in units of the mean splitting, varies monotonically from $0.32\bar{\Omega}_1/\bar{\Omega}_0$ to $0.40\bar{\Omega}_1/\bar{\Omega}_0$ as l varies from 4 to 10.

Consider first multiplets 4–7 in Table 1, whose kernels are concentrated near the surface of the Sun. The mean σ is about 0.015 of the mean splitting, which implies, if the modes have been identified correctly, that $\bar{\Omega}_1/\bar{\Omega}_0 \leq 0.04$. This is to be compared with the differential rotation in the photosphere, which satisfies $\Omega_1(R)/\Omega_0(R) \approx 0.2$. The result implies either that the θ -dependence of Ω is confined to a boundary layer about $0.05R$ deep, or that Ω_1 changes sign, so that at the base of the convection zone the rotation is faster at the poles than at the equator.

Note that among the surface modes, multiplets 6 and 7 are the more concentrated near the surface. Therefore one would expect their rotational splitting to be the less uniform, which is apparently not the case. Aside from the possibility that the data have been wrongly interpreted, this suggests either that the magnetic field increases with depth or that the modes have been incorrectly identified. According to equation (13), nonuniformity in the splitting increases as m increases, so it is quite conceivable that the components with the greatest values of m have not been recognized. The multiplet 6 could be a high-

degree f mode, such as $f(l=32)$, and multiplet 7 could be $p_2(l=10)$. It is unlikely that multiplet 6 is a high-degree g mode, even though there are such modes with appropriate frequencies, because these modes are likely to be severely trapped beneath the convection zone and hence be unobservable^{31,32}.

Multiplets 1–3 in Table 1 are concentrated in the core. The high degree of uniformity observed in the splitting implies that the core's rotation is nearly uniform on spheres. The limits set by σ yield $\bar{\Omega}_1/\bar{\Omega}_0 \leq 7 \times 10^{-3}$. They also set limits on the magnetic field in the core. If a field of structure similar to that adopted by Dicke²⁷ is assumed, the mean field intensity is bounded above by about 3×10^4 G.

Radial variation of Ω

Because the contribution to ν_{nlm}' from Ω_1 is small, I shall neglect it. Then the mean observed splittings Δ listed in Table 1 yield seven differently weighted averages of Ω_0 . These are

$$\int_0^R RK_{nl}(r)\Omega_0(r)dr/R = \beta_{nl}^{-1}(2\pi\Delta + \omega_e) \quad (14)$$

where ω_e is the orbital angular velocity of the Earth. The values of these averages are listed in Table 1.

Provided the data are not mutually inconsistent, inversion of equations (14) to obtain a plausible estimate of Ω_0 is possible, but not unique. Nonuniqueness arises partly because there are only a finite number of averages to constrain the function. Therefore one must adopt some criterion for preference. Here I follow an optimal averaging procedure of Backus and Gilbert³³. For each value r_0 of r one attempts to construct a linear combination $D(r, r_0) = \sum \alpha_{nl}(r_0)K_{nl}(r)$ of the data kernels K_{nl} in such a way that D is concentrated near r_0 . The α_{nl} are normalized to make D unimodal. The average $W(\bar{r})$ of Ω_0 weighted with the new kernel D is simply $\sum \alpha_{nl}\beta_{nl}^{-1}(2\pi\Delta + \omega_e)$. Here $\bar{r}$ is the median of D . If the attempt to concentrate D was successful, then D resembles a delta function, and one may safely set $\Omega_0(r) \approx W(r)$. If D is not well concentrated, one may still adopt $W(r)$ as the least biased estimate of $\Omega_0(r)$, but one must bear in mind that structure on a scale smaller than the width of D cannot be resolved. Thus the function W is likely to be smoother than Ω_0 . Moreover, it does not satisfy the constraints (14) precisely. Note, however, that W is an invariant property of data, in the sense that all functions $\Omega_0(r)$ that satisfy the constraints (14) have the same averages $W(\bar{r})$ when weighted with the averaging kernels D .

The Backus–Gilbert inversion procedure has been applied to the data listed in Table 1. In so doing, a further constraint of the form (14) was added, with a data kernel that is a delta function at $r=R$ and a right-hand side equal to $2\pi/(25.4 \text{ days}) = 2.86 \times 10^{-6} \text{ s}^{-1} \equiv \Omega_s$. This device ensures that $\Omega(R) = \Omega_s$. The function $W(r)$ is shown in Fig. 2.

To give an idea of the extent of the averaging, the half-width δ of D , defined as $r_2 - r_1$ where

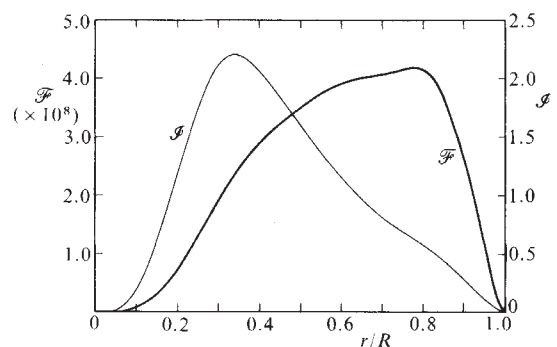


Fig. 4 The moment of inertia kernel $\mathcal{F}$ and the oblateness kernel $\mathcal{P}$, extrapolated linearly from the corresponding kernels of models A and B of ref. 30 as in equation (17) with $\lambda = 1.26$. $\mathcal{F}$ is measured in units of the moment of inertia I , and $\mathcal{P}$ is measured in units of Ω_s^{-2} .

$$\int_{r_1}^F D dr = \int_F^{r_2} D dr = \frac{1}{4} \quad (15)$$

is also displayed. The averages are quite well confined near the centre and the surface, but are very broad in the mid-range. This is not surprising, as none of the data kernels K_{nl} has a substantial amplitude where $0.2 \leq r/R \leq 0.8$. The failure of the Backus–Gilbert procedure to provide a kernel D that is well concentrated about any point in the mid-range results from a paucity of appropriate data, and does not reflect a deficiency in the procedure. Though $\Omega_0(r)$ is quite strongly constrained near the centre and the surface, there is a wide margin of uncertainty elsewhere.

An important goal of the inversion procedure is the attainment of a stable solution. If averaging kernels D are chosen so as to minimize δ , small changes to either the data or the kernels K_{nl} yield large changes in W when several different data, or linear combinations of them, provide nearly the same physical information. Thus errors are magnified and the solution is inaccurate. Backus and Gilbert³³ discuss how, at the expense of permitting somewhat broader averaging kernels D , a tradeoff between stability to errors in the data (or the mapping of those data on to Ω_0 via the data kernels K_{nl}) and formal mathematical accuracy can be achieved. That procedure has been adopted here. The rotation curve presented in Fig. 2 is stable to small changes in the data or the kernels, and is therefore more likely to be a good representation of the actual angular velocity of the Sun than is a curve constructed to fit the constraints (14) exactly.

As an indication of the uncertainty in $\Omega_0(r)$ I display four alternative possibilities in Fig. 3. The first is the fourth-degree polynomial in r satisfying $\Omega_0(R) = \Omega_s$ that best fits equations (14) by least squares. Its shape is quite sensitive to changes in the equilibrium solar model. (The corresponding seventh-degree polynomial, which is even more sensitive to the equilibrium model, satisfies equations (14) exactly. Some of its properties are listed in Table 2.) The second is in some sense the most slowly varying function that fits the data, for it minimizes the integral of $(d\Omega_0/dr)^2$ among all differentiable functions Ω_0 that satisfy the constraints (14) and $\Omega_0(R) = \Omega_s$. The third is of the form $\Omega_0 = \tilde{\Omega} = \text{constant}$ for $0 \leq r < r_c$ and $\Omega_0 = (R/r)^2 \Omega_s$ for $r_c < r \leq R$. This function was chosen, somewhat arbitrarily, with the idea that the radiative interior might be rotating rigidly, aside from a thin boundary layer at the base of the convection zone. Because Ω_0 is very much smaller than the buoyancy frequency in the radiative zone, only a slight smoothing of the discontinuity would be necessary to satisfy the Solberg–Høiland dynamical stability criterion^{34–36} in the equatorial plane, with negligible effect on the values of the angular momentum, rotational kinetic energy and J_2 . Again the free parameters were chosen to give the closest fit to equation (14) by least squares, yielding $r_c = 0.69R$ and $\tilde{\Omega} = 1.86 \times 10^{-5} \text{ s}^{-1}$. The value of r_c is close to independent modern estimates^{37–40} of the radius of the base of the convection zone. The fourth function is of the form $\Omega_0 = \tilde{\Omega} [1 - \alpha(r/R)^b]$ for $0 \leq r \leq r_c$ and $\Omega_0 = (R/r)^2 \Omega_s$ for $r_c \leq r \leq R$, with $\tilde{\Omega}$, α , b , r_c chosen to give the least squares fit to equation (14) subject to Ω_0 being continuous.

Angular momentum, rotational kinetic energy and J_2

From each of the rotation curves displayed in Figs 2 and 3 it is straightforward to compute the angular momentum H , the rotational kinetic energy T and the measure J_2 of the gravitational quadrupole moment. These are given by

$$H = \int_0^1 \mathcal{J} \Omega_0 dx, \quad T = \frac{1}{2} \int_0^1 \mathcal{J} \Omega_0^2 dx, \quad J_2 = \int_0^1 \mathcal{F} \Omega_0^2 dx \quad (16)$$

where $x = r/R$, $\mathcal{J}(x) = (8\pi/3)R^5 \rho x^4$ and $\mathcal{F}$ is given in ref. 17. The results are listed in Table 2. The kernels $\mathcal{J}$ and $\mathcal{F}$ are displayed in Fig. 3. Notice that they are greatest where knowledge of Ω_0 is the least. This unfortunate property magnifies the uncertainty in H , T and J_2 .

On the solar helium abundance

A least squares fit of the frequencies

$$\nu_\lambda = \lambda \nu_A + (1 - \lambda) \nu_B \quad (17)$$

interpolated or extrapolated linearly from the eigenfrequencies ν_A and ν_B of models A and B of ref. 30 (which have initial helium abundances $Y = 0.25, 0.19$ respectively) to observed frequencies of low-degree 5-min modes³⁸ yields $\lambda = 1.36 \pm 0.12$. Thus it seems that Y is somewhat greater than 0.25. A similar fit to the frequencies in Table 1 yields $\lambda = 1.26 \pm 0.79$. The latter fit is hardly strong independent evidence that $Y > 0.25$, especially since the mode identifications are uncertain. Nevertheless, it does show that the data in Table 1 are apparently consistent with this result. All the computations listed in this paper used the extrapolated model with $\lambda = 1.26$. Using model A typically changes H^2 , T and J_2 by no more than 10%.

Conclusion

If the sequence of peaks found in the spectra of Bos and Hill²² really are a consequence of rotational splitting, then it may confidently be concluded that the central core of the Sun rotates on average some six times faster than the surface. Moreover, the angular velocity Ω in the core is almost independent of latitude. The distribution of angular velocity throughout the rest of the Sun is uncertain, but it is not unlikely to resemble the solid curve in Fig. 2. Of course, there may be variations about that curve on a scale less than δ (shown also in Fig. 2). Such structure cannot possibly be resolved by observations of only the modes listed in Table 1, however precise the measurements. What is needed is to observe modes with data kernels that have substantial amplitudes where $0.2 \leq r/R \leq 0.8$.

Aside from a smoothed version of curve 3 in Fig. 3, none of the estimates of Ω_0 is dynamically stable. All the estimates are unstable to the Rayleigh criterion $d(\omega^2 \Omega_0^2)/d\omega > 0$, where ω is the distance from the rotation axis. This condition is relevant when thermal diffusion is taken into account^{41–43}. Indeed, it is interesting that there exists no function Ω_0 satisfying $\Omega_0(R) = \Omega_s$ and the Rayleigh criterion that can reproduce the frequency splitting of multiplets 4–7 in Table 1. Therefore the data seem to imply either that the mean flow is stabilized by some agent, such as a magnetic field, or that the natural tendency of the instability to restore Ω_0 to a state satisfying the Rayleigh criterion is effectively countered by some other angular-momentum transporting process.

The splitting data alone cannot determine Ω_0 . Thus, to arrive at an estimate it is necessary to impose at least one additional restriction. That chosen here is one of smoothness, and was applied by equating the solid curve in Fig. 2, which is a well-determined average of Ω_0 with a known kernel D , with Ω_0 itself. This procedure has been applied successfully in other inverse problems, particularly in geophysics. The result is likely to be smoother than the actual angular velocity.

The curve depicted in Fig. 2 implies that the measure J_2 of the gravitational quadrupole moment is 3.6×10^{-6} . It is difficult to judge how accurate this value is, partly because the identification of the modes is uncertain, and partly because even if the modes were known, the constraints imposed by the splitting alone do not restrain the possible functions Ω_0 within tight bounds. The smallest value of J_2 compatible with the data is about 1.2×10^{-6} ; the data cannot bound J_2 above.

The implications the estimate has on the test of gravitation theories using the precession of the orbit of Mercury can be assessed from equations (5)–(7). Taking $J_2 = 3.6 \times 10^{-6}$ yields

$$\frac{1}{3}(2 - \beta + 2\gamma) = \begin{matrix} 0.992 \pm 0.005 & (\text{ref. 6}) \\ 0.996 \pm 0.005 & (\text{ref. 3}) \end{matrix} \quad (18)$$

The errors quoted are standard errors arising from the orbit analysis, and take no account of the uncertainty in J_2 .

Finally, it should be remembered that a careful analysis of the power spectra has yet to be undertaken, and that the uniformly spaced sequences of peaks might have been interpreted wrongly.

I thank Professor H. A. Hill for collaboration in Cambridge last November when together we isolated some of the multiplets in the power spectrum, identified the modes of oscillation, constructed a crude rotation curve and made our first estimate of J_2 . Since then our opinions about the implications of the

data have diverged, and we have decided to publish our findings separately (see ref. 44). I thank J. Christensen-Dalsgaard for some computer programs and comments, and D. Gubbins, D. Lynden-Bell and K. Whaler for discussions, also J. E. Pringle for suggesting improvements on an early version of the paper.

Received 13 April; accepted 4 June 1982.

1. Misner, C., Thorne, K. S. & Wheeler, J. A. *Gravitation* (Freeman, San Francisco, 1973).
2. Will, C. M. *General Relativity: An Einstein Centenary Survey* (eds Hawkins, S. W. & Israel, W.) 24–89 (Cambridge University Press, 1979).
3. Anderson, J. D., Keese, M. S. W., Lau, E. L., Standish, E. M. Jr & Newhall, xx *Astronautica* 5, 43–61 (1978).
4. Dicke, R. H. & Goldenberg, H. M. *Astrophys. J. Suppl.* 27, 131–182 (1974).
5. Hill, H. A. & Stebbins, R. T. *Astrophys. J.* 200, 471–483 (1975).
6. Shapiro, I. I., Counselman, C. C. III & King, R. W. *Phys. Rev. Lett.* 36, 555–558 (1976).
7. Anderson, J. D., Colombo, G., Friedman, L. D. & Lau, E. L. *Proc. Int. Meet. on Experimental Gravitation* (ed. Bertotti, B.) 393–422 (Accademia Nazionale dei Lincei, Rome, 1977).
8. Esterbrook, F. B. et al. *Sources of Gravitational Radiation* (ed. Smarr, L. L.) 37–48 (Cambridge University Press).
9. Christensen-Dalsgaard, J. & Gough, D. O. *Nonradial and Nonlinear Stellar Pulsations* (eds Hill, H. A. & Dziembowski, W. A.) 369–380 (Springer, Heidelberg, 1980).
10. Nordvedt, K. Jr *Proc. int. Meet. on Experimental Gravitation* (ed. Bertotti, B.) 247–256 (Accademia Nazionale dei Lincei, Rome, 1977).
11. Hill, H. A. *The New Solar Physics* (ed. Eddy, J. A.) 135–214 (Westview, Boulder, 1978).
12. Severny, A. B., Kotov, V. A. & Tsap, T. T. *Nature* 259, 87–89 (1976).
13. Brookes, J. R., Isaak, G. R. & van der Raay, H. B. *Nature* 259, 92–95 (1976).
14. Christensen-Dalsgaard, J. & Gough, D. O. *Nature* 259, 89–92 (1976).
15. Cowling, T. G. & Newing, R. A. *Astrophys. J.* 109, 149–158 (1949).
16. Ledoux, P. *Astrophys. J.* 144, 373–384 (1951).
17. Lynden-Bell, D. & Ostriker, J. P. *Mon. Not. R. astr. Soc.* 136, 293–310 (1967).
18. Hansen, C. J., Cox, J. P. & Van Horn, J. M. *Astrophys. J.* 217, 151–159 (1977).
19. Cuyper, J. A. *Astrophys. J.* 89, 207–208 (1980).
20. Gough, D. O. *Mon. Not. R. astr. Soc.* 196, 731–745 (1981).
21. Scott, S. *The Application of Modern Physics to the Earth and Planetary Interiors* (ed. Runcorn, S. K.) 586–602 (Wiley-Interscience, London, 1969).
22. Bos, R. J. & Hill, H. A. *Solar Phys.* (in the press).
23. Hill, H. A., Bos, R. J. & Caudell, T. P. *Solar Phys.* (in the press).
24. Hill, H. A. *Proc. R. astr. Soc. Meet. in Dublin, Irish astr. J.* (in the press).
25. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* 293, 443–445 (1981).
26. Isaak, G. R. *Nature* 296, 130–131 (1982).
27. Dicke, R. H. *Proc. Informal Conf. on Status and Future of Solar Neutrino Research* Vol. 2 (ed. Friedlander, G.) 109–143 (Brookhaven National Laboratory, Upton, 1978).
28. Dicke, R. H. *New Scientist* 83, 12–14 (1979).
29. Gough, D. O. *Nature* 298, 350–354 (1982).
30. Christensen-Dalsgaard, J., Gough, D. O. & Morgan, J. G. *Astr. Astrophys.* 73, 121–128; 79, 260 (1979).
31. Dziembowski, W. A. & Pamjatnykh, A. A. *Pleins Feux sur la Physique Solaire* (ed. Rösch, J.) 135–140 (CNRS, Paris, 1978).
32. Christensen-Dalsgaard, J., Dziembowski, W. A. & Gough, D. O. *Nonradial and Nonlinear Stellar Pulsation* (eds Hill, H. A. & Dziembowski, W. A.) 313–341 (Springer, Heidelberg, 1980).
33. Backus, G. & Gilbert, F. *Phil. Trans. R. Soc. A266*, 123–192 (1970).
34. Solberg, H. *Procès-verbaux de l'assn. de météor. UGGI 6^e Assemblée Générale (Edinburgh)*, *Mém. et Disc.* 2, 66–82 (1936).
35. Høiland, E. *Avhandl. Utgitt Norske Vidensk.—Akad. Oslo I. Mat.—naturv. Kl. No. 11*, 1–24 (1941).
36. Zahn, J.-P. *IAU Symp. No. 59*, 185–195 (1974).
37. Lubow, S. H., Rhodes, E. J. Jr & Ulrich, R. K. *Nonradial and Nonlinear Stellar Pulsation* (eds Hill, H. A. & Dziembowski, W. A.) 300–306 (Springer, Heidelberg, 1980).
38. Berthomieu, G. et al. *Nonradial and Nonlinear Stellar Pulsation* (eds Hill, H. A. & Dziembowski, W. A.) 307–312 (Springer, Heidelberg, 1980).
39. Christensen-Dalsgaard, J. & Gough, D. O. *Astr. Astrophys.* 104, 173–176 (1981).
40. Scuflaire, R., Gabriel, M. & Noels, A. *Astr. Astrophys.* 99, 39–42 (1981).
41. Yih, C.-S. *Phys. Fluids* 4, 806–811 (1961).
42. Goldreich, P. & Schubert, G. *Astrophys. J.* 150, 571–587 (1967).
43. Fricke, K. J. *Z. Astrophys.* 68, 317–344 (1968).
44. Hill, H. A., Bos, R. & Goode, P. R. *Phys. Rev. Lett.* (submitted).

The thermal springs of Bath

J. N. Andrews*, W. G. Burgess†, W. M. Edmunds†, R. L. F. Kay‡ & D. J. Lee§

* School of Chemistry, University of Bath, Bath BA2 7AY, UK

† Institute of Geological Sciences, Hydrogeology Unit, Wallingford, Oxon OX10 8BB, UK

‡ Institute of Geological Sciences, Environmental Protection Unit, Harwell, Didcot, Oxon OX11 0RA, UK

§ United Kingdom Atomic Energy Authority, Winfrith, Dorset DT2 8DH, UK

Geochemical and hydrogeological evidence shows that the thermal springs at Bath originate from the Mendip Hills. A maximum subsurface temperature of $80 \pm 16^\circ\text{C}$ is attained during circulation to between 2.7 and 4.3 km within the Carboniferous Limestone. The residence time of the bulk of the water is $< 10,000$ yr.

INTEREST in the evaluation of the geothermal resources of Britain has prompted a thorough geochemical investigation of the Bath thermal springs and their regional setting¹. The King's Spring, the principal spring at Bath, emerges through Lower Lias shales within a reservoir of Roman construction. It has a flow of 13 l s^{-1} at a temperature of 46.5°C which records show to have been constant within $\pm 0.5^\circ\text{C}$ since 1754. Its thermal yield above a non-thermal groundwater temperature of 10°C is 2.0 MW. Two smaller springs at Bath, the Cross Bath and Old Royal springs, have chemical compositions very similar to that of the King's Spring. Thermal water also occurs at Bristol where the Hotwells Spring discharges from Carboniferous Limestone into river muds just above low tide level in the Avon gorge. The flow is about 0.2 l s^{-1} at 24°C and the thermal yield is only 7 kW.

The Bath/Bristol district is a complex synclinal structure formed between three intersecting fold axes². The most important aquifers which might contribute to the flow of the thermal water, are in the Carboniferous Limestone and in Triassic deposits. The Old Red Sandstone (Devonian) and the Pennant Sandstones (Carboniferous) have only restricted hydraulic continuity.

The origin and composition of the Bath thermal waters have been the subject of much scientific curiosity. Notable associated discoveries include the extraction of helium from the exsolved gases by Sir James Dewar³; the identification of their radio-

activity due to radium by Strutt³ in 1904 and that due to radon by Munro⁴ in 1928. However, no acceptable explanation of the heat source and circulation history of the waters has yet been proposed. The objective of the present study was to use a range of geochemical techniques to deduce the source, the circulation history, the temperature at depth and the age of groundwater⁵.

Geochemical characteristics

(1) Inorganic chemistry: the King's Spring water has a total mineralization of $2,150\text{ mg l}^{-1}$ and Ca^{2+} and SO_4^{2-} are the dominant ions (Table 1). The chemical composition remained constant over the 2 yr period of investigation, and comparison with earlier work^{6–9} suggests that the composition is stable over a long period. This stability is also reflected in the fact that the thermal water is in equilibrium with the minerals calcite, dolomite, gypsum, fluorite and barite as calculated using the WATEQF program¹⁰.

Regional sampling of representative groundwaters from local aquifers was carried out and these results⁵ show that no other groundwaters with a composition similar to that of Bath thermal water are found in the region, except locally in the Triassic aquifer, as discussed below. The Hotwells spring at Bristol has a composition equivalent to a mixture of Bath-type thermal water and shallow, Carboniferous Limestone water in the ratio 1:2.3.

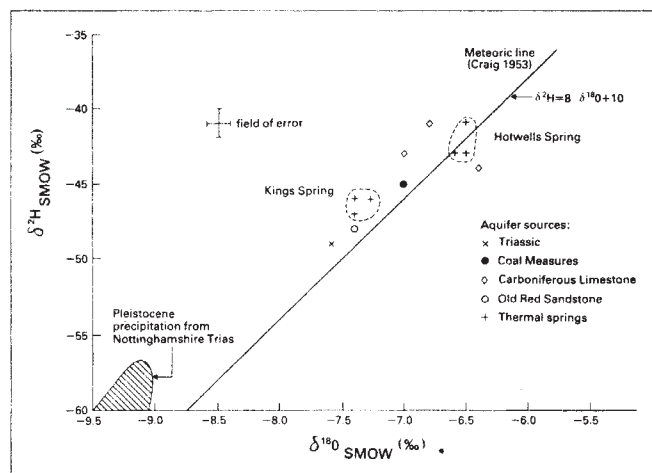


Fig. 1 Plot of hydrogen and oxygen stable isotope results for the thermal waters in relation to other groundwaters in the region.

(2) Dissolved radioelements: The isotope ^{234}U can be dissolved in preference to its precursor, ^{238}U , as a consequence of its genesis by decay processes and the $^{234}\text{U}/^{238}\text{U}$ activity ratio in groundwaters is frequently in excess of equilibrium¹¹. The Bath thermal water is distinguished from other groundwaters in the region by a very low U-content and a high $^{234}\text{U}/^{238}\text{U}$ activity ratio (Table 1c). Only groundwaters from the Old Red Sandstone have similar activity ratios although they generally have higher U-contents than thermal water.

The high level of radioactivity in the King's Spring water is due to ^{222}Rn and during 1977 it gradually increased from 1,000 to 2,220 pCi kg⁻¹, the value determined¹² in 1972. This increase is attributed to a delayed piezometric recovery following the unusual drought conditions of 1976. Other radiochemical parameters and the radiogenic ^4He content similarly increased during 1977. ^{222}Rn contents similar to that of the King's Spring also occur locally in groundwaters from the Trias. Although the ^{226}Ra activity in the water is much lower than that of ^{222}Rn , it is the most significant radioelement present with regard to radiological hazard¹³. The short-lived β -emitting ^{222}Rn daughters, $^{214}\text{Pb}/^{214}\text{Bi}$ are present in equilibrium with ^{222}Rn . The isotope ^{210}Pb (half life 22 yr) requires about 100 yr to reach equilibrium with a constant ^{222}Rn source. Because the activity of ^{210}Pb in the thermal water is negligible although there is adequate natural Pb carrier present, this suggests either that the ^{222}Rn content has not been constant for a long period or that rapid wall-rock exchange of lead isotopes must be occurring.

The natural Th content of the King's Spring is below the detection limit of 0.002 μg per kg. The β -emitting ^{232}Th daughters, $^{212}\text{Pb}/^{212}\text{Bi}$, are present to the extent of 2.6 pCi kg⁻¹. These isotopes must be derived by ^{220}Rn release into the groundwater by mechanisms similar to those for ^{222}Rn release. The ratio of $^{222}\text{Rn}/^{220}\text{Rn}$ activities suggests that the U/Th ratio in the source rock could be as high as 250:1, though this may be an overestimate as the efficiency of ^{220}Rn release could be much less than that for ^{222}Rn . However, the uranium geochemistry also suggests that U deposition has occurred somewhere in the flow system.

(3) Dissolved gases: The gas composition (Table 1d) can be explained as the exsolution of atmospheric gases, which were dissolved at recharge, modified by the geochemical reaction of oxygen in the aquifer, by the addition of radiogenic ^4He due to radioelement decay and by the addition of hydrocarbon gases. At atmospheric pressure, the ratio of nitrogen solubility at 46.5°C to that at 9.5°C, the approximate recharge temperature, is 0.6, so that less than half of the dissolved gas should be exsolved as the thermal water re-equilibrates with the atmosphere. The total dissolved gas flow was calculated as 710 l h⁻¹ in the present study and this may be compared with estimates¹⁴ of the total exsolved gas flow of 208 l h⁻¹. Seasonal fluctuations

Table 1 The composition of thermal water from the King's Spring, Bath

a, Major element composition (mg kg⁻¹)

Year source	1977-79 (this study)	1961 (ref. 6)	1936 (ref. 7)
	Range	Mean	
Temperature (°C)	46.4-46.6	46.5	48
Flow (ls ⁻¹)	—	15.0	—
pH	6.39-6.94	6.68	7.1
Eh (mV)	+220-32	—	—
O ₂ (mg l ⁻¹)	0.5-6.0	1.8	—
Total solids	2,177-2,280	2,203	2,166
Na ⁺	169-190	183	174
K ⁺	16.2-20.3	17.4	15.7
Ca ²⁺	352-399	382	392
Mg ²⁺	52.0-55.4	53	54
HCO ₃ ⁻	184-206	192	216
SO ₄ ⁻	1,010-1,060	1,032	1,021
Cl ⁻	280-292	287	276

		1888 (ref. 8)	1874 (ref. 9)
Temperature (°C)	46.4-46.6	46.5	47
Flow (ls ⁻¹)	—	15.0	—
pH	6.39-6.94	6.68	—
Eh (mV)	+220-32	—	—
O ₂ (mg l ⁻¹)	0.5-6.0	1.8	—
Total solids	2,177-2,280	2,203	2,360
Na ⁺	169-190	183	135
K ⁺	16.2-20.3	17.4	31
Ca ²⁺	352-399	382	402
Mg ²⁺	52.0-55.4	53	52
HCO ₃ ⁻	184-206	192	88
SO ₄ ⁻	1,010-1,060	1,032	1,061
Cl ⁻	280-292	287	277

b, Trace element composition (mg kg⁻¹)

		1961 (ref. 6)	1936 (ref. 7)
Li ⁺	0.240-0.255	0.242	0.050
Sr ²⁺	5.60-6.10	5.92	8.0
Ba ²⁺	—	0.024	0.8
Total Mn	0.065-0.085	0.068	0.059
Total Fe	0.950-1.320	0.879	0.009
Ni ²⁺	0.002-0.030	0.022	0.0005
Cu ²⁺	0.0005-0.0037	0.0021	0.0035
F ⁻	2.0-2.20	2.6	2.0
Br ⁻	1.00-1.50	1.32	6.9
NO ₃ ⁻	0.9-1.7	1.2	<0.0002
HPO ₄ ²⁻	—	0.020	0.0058
SiO ₂	43.2-44.3	44.1	44.1

c, Isotopic and radioelement composition (this study)

δ ¹⁸ O	‰ rel. SMOW	-7.4
δ ² H	‰ rel. SMOW	-47.0
δ ³ H	Tritium units	0-3.1
	pCi kg ⁻¹	0-9.0
δ ¹³ C	‰ rel. PDB	-1.5
δ ¹⁴ C	% modern carbon	4.5

U-series isotopes	Mode	1977-79 Range
Natural U	$\mu\text{g kg}^{-1}$	0.05
^{238}U	pCi kg ⁻¹	0.017
$^{234}\text{U}/^{238}\text{U}$	activity ratio	2.5
^{226}Ra	pCi kg ⁻¹	11.0
^{222}Rn	pCi kg ⁻¹	2,350
$^{214}\text{Pb}/^{214}\text{Bi}$	pCi kg ⁻¹	2,367 ± 200
^{210}Pb	pCi kg ⁻¹	<1.0

Th series isotopes

Natural Th	$\mu\text{g kg}^{-1}$	<0.002
$^{212}\text{Pb}/^{212}\text{Bi}$	pCi kg ⁻¹	2.61 ± 0.26
$^{224}\text{Ra}/^{220}\text{Rn}$	pCi kg ⁻¹	2.61 ± 0.26

d, Exsolved gas composition and dissolved noble gas content

Exsolved gas composition %		Dissolved noble gases cm ³ STP/cm ³ H ₂ O
N ₂	to 100%	^4He 1,080 × 10 ⁻⁸
O ₂	<0.5	Ne 2.10 × 10 ⁻⁷
^4He	0.1	Ar 4.04 × 10 ⁻⁴
Ar	1.0	Kr 0.91 × 10 ⁻⁷
CO ₂	3.5	Xe 1.19 × 10 ⁻⁸
CH ₄	0.6	
C ₃ H ₈	0.8	

in dissolved oxygen ($<0.5\text{--}6.0\text{ mg l}^{-1}$) and in Eh (redox potential) were also noted and may indicate some mixing of waters at intermediate depths.

Source of thermal groundwater and its recharge

The hydrogen and oxygen isotopic compositions of the thermal water demonstrate that it is of meteoric origin (Fig. 1). The isotopic compositions lie close to the worldwide meteoric line¹⁵ and the ratios are similar to those of modern shallow groundwaters in the region. The Bath thermal water may be clearly distinguished from known occurrences¹⁶ of late Pleistocene groundwater elsewhere in England where a recharge temperature some 5°C cooler than postglacial climatic conditions has been indicated. The ratios of dissolved atmospheric inert gases in the thermal waters indicate recharge temperatures around 9°C . The combined evidence shows that climatic conditions at recharge were similar to those of the present day.

The piezometric levels in the various aquifers in the Bath/Bristol region are shown in Fig. 2. The piezometric level of the King's Spring is known from the historical record for a well which encountered thermal water in Kingsmead Street¹⁷ in 1836. It is 40 m above sea level (10 m above the elevation of the spring). The only location with sufficient head to cause flow at the King's Spring is the Mendip Hills, about 15 km to the south-west and on the margin of the basin.

Maximum temperature and circulation depth

Rock-water equilibration temperatures for groundwaters from geothermal areas have been estimated from the amounts of dissolved Na^+ , K^+ , Ca^{2+} and SiO_2 (refs 18, 19). It was found that the cation geothermometers were not valid for the sedimentary formations of the Bath/Bristol region. The application of the SiO_2 geothermometer depends on which silica polymorph controls its dissolution. The thermal groundwaters are closer to equilibrium with chalcedony ($\text{SI} = +0.18$) than with quartz ($\text{SI} = +0.60$) but because of the limited number of springs the control on SiO_2 solution is uncertain. Maximum rock equilibration temperatures of 64°C (for chalcedony control) and 96°C (for quartz control) have been calculated and the maximum temperature reached at depth probably lies within this range. The SiO_2 geothermometer indicates that the thermal component of the Hotwells source has been equilibrated in the range $49\text{--}72^\circ\text{C}$ before subsequent mixing with shallow groundwater from the Carboniferous Limestone.

The necessary circulation depth for meteoric water to reach these temperatures may be estimated from the geothermal gradient. The only geothermal gradient measurement²⁰ (11°C km^{-1}) in the basin, however, is known to have been affected by cold water inflows in the borehole. On the assumption that the region is one of average heat flow (40 mW m^{-2}), a geothermal gradient close to 20°C km^{-1} is more likely (J. R. Bloomer, personal communication). This would place the circulation depth of the Bath thermal water in the range 2,700–4,300 m to attain a temperature of $64\text{--}96^\circ\text{C}$ and this depth strongly implies that either the Carboniferous Limestone and/or the Old Red Sandstone must be the principal storage aquifer(s).

The slightly negative $\delta^{13}\text{C}$ (-1.5%) of the dissolved bicarbonate is very strong evidence that the thermal water has equilibrated chemically with the marine Carboniferous Limestone rather than with the largely non-marine Old Red Sandstone. The value of $\delta^{13}\text{C}$ for the latter²¹ is -10.7% and is too negative to permit the development of the observed $\delta^{13}\text{C}$ of the thermal water. It must be concluded that the most recent storage and chemical equilibration of the carbonate system takes place within the Carboniferous Limestone ($\delta^{13}\text{C} = -0.05\%$), although a proportion of the thermal water may previously have been transmitted in the Old Red Sandstone.

Thermal water age

Tritium, ^3H , produced by thermonuclear weapon testing, was used to identify any post-1950s recharge. The ^3H content of

the King's Spring averaged $1 \pm 2\text{ TU}$ over the monitoring period and the highest observed value was $3.1 \pm 2\text{ TU}$. This, together with the small amounts of nitrate present, variations in dissolved O_2 , Eh and $^{234}\text{U}/^{238}\text{U}$ activity ratio, suggests that there could be some mixing of the thermal water with very small quantities of 'recent' waters from shallow aquifers.

The ^{14}C specific activity for the dissolved bicarbonate of the King's Spring averages 4.5% of that for modern carbon (Table 1c). The uncertain source and age of any 'young' mixing component leads to uncertainty in the interpretation of the carbon isotope data in terms of residence time for the thermal water. Assumptions about the extent of mixing, however, allow limiting conditions to be defined. If there is no mixing with recent groundwater, and the ^{14}C activity represents that remaining after decay of the original ^{14}C present, a corrected age for the thermal water of 20,000 yr may be derived on the basis of the stoichiometric reaction between biogenic CO_2 (100% modern carbon; $\delta^{13}\text{C} = -26.1\%$) and ^{14}C -dead carbonate in the recharge zone (no ^{14}C ; $\delta^{13}\text{C} = -0.05\%$ for Carboniferous Limestone). Such a 1:1 reaction between carbonate and biogenic CO_2 , however, would result in a $\delta^{13}\text{C}$ value of -13.0% for the dissolved bicarbonate.

The measured $\delta^{13}\text{C}$ of the thermal water bicarbonate (Table 1c) is -1.6% , and this indicates that extensive bicarbonate/carbonate interphase exchange has occurred after the initial congruent solution of carbonate. It is not possible to estimate, or correct for, the concurrent loss of ^{14}C from the bicarbonate in solution to the aquifer carbonate in this process. The age estimate of 20,000 yr is therefore a maximum possible age, provided that no ^{14}C activity has been introduced by mixing with shallow groundwater of recent origin. Alternatively, if mixing with recent groundwater (which may be dead to ^3H , but contain appreciable ^{14}C) is sufficient to account for all the observed ^{14}C activity, then the thermal water itself is dead to ^{14}C and is older than 50,000 yr.

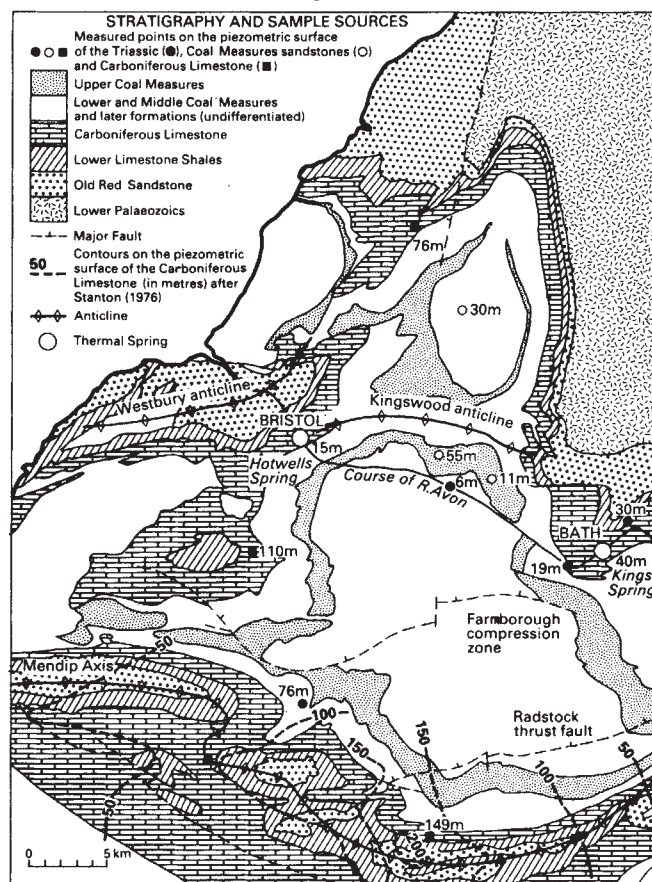


Fig. 2 Outcrop and subcrop of Palaeozoic strata in the Bath-Bristol region, with major structural features (after G. A. Kellaway, unpublished data); the localities sampled during the study and the piezometric data are also shown.

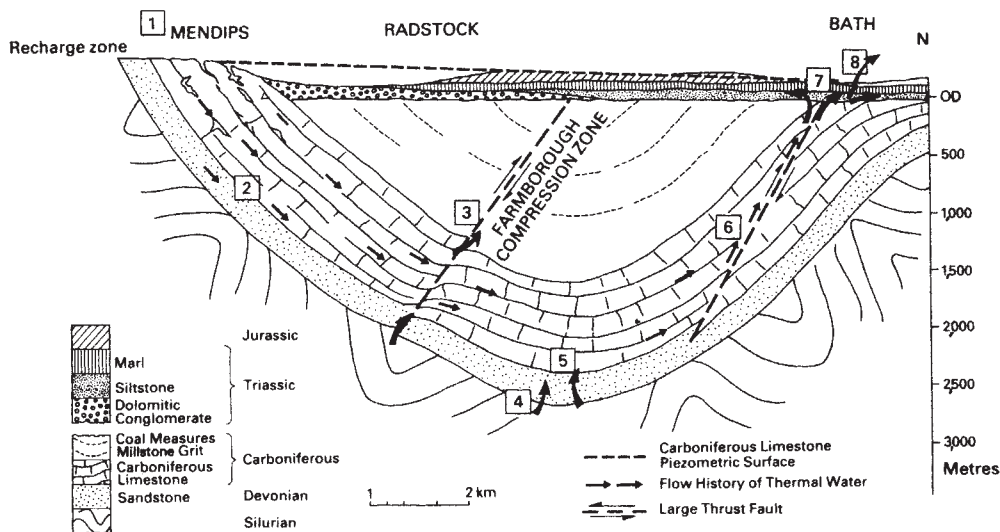


Fig. 3 Conceptual model of the flow path of thermal water in the Bath-Bristol Basin: 1, recharge (9–10 °C) at the Carboniferous Limestone/Old Red Sandstone outcrop; 2, down dip and down gradient flow, assisted by karstic development; 3, possible gain from Lower Palaeozoics and leakage to Upper Carboniferous via Farmborough compression zone; 4, some leakage of very old ^4He -bearing groundwater from Old Red Sandstone and possibly Lower Palaeozoics; 5, storage and chemical equilibration within the Carboniferous Limestone, at 64–96 °C; 6, relatively rapid ascent along thrust fault system; 7, some recharge of Triassic by thermal water at Bath; 8, discharge of the thermal springs at Bath (46.5 °C).

The temperature at which the thermal water was recharged (9–10 °C), estimated from its inert gas contents and its isotopic composition ($\delta^2\text{H}$ and $\delta^{18}\text{O}$), shows that recharge took place in climatic conditions similar to those of the Holocene. This implies either that the water is younger than 10,000 yr (the beginning of the present climatic optimum), or older than 70,000 yr (the end of the last interglacial). It is therefore not possible to distinguish between the alternative explanations for the ^{14}C activity in the water, on the basis of its recharge temperature.

The ^4He content of a groundwater increases with its residence time in a formation of porosity, ϕ , and density, ρ , according to the equation²²:

$$^4\text{He content, cm}^3/\text{cm}^3 \text{ H}_2\text{O} = \rho \phi^{-1} \{1.19 \times 10^{-13} [\text{U}] + 2.88 \times 10^{-14} [\text{Th}]\}t \quad (2)$$

where [U] and [Th] are the U and Th contents of the aquifer rock and t is the groundwater age in years if all the ^4He generated is dissolved in the pore fluids. This equation is not strictly applicable to the Carboniferous Limestone which is a fracture-flow system for which estimation of effective porosity is difficult. It is also unlikely that all the ^4He generated would be dissolved in the fracture fluids if the porosity is low. However, equation (2) yields a minimum age of 250,000 yr for a porosity of 1% and the radioelement contents of the Carboniferous Limestone. The Old Red Sandstone has radioelement contents about four times higher than those of the Carboniferous Limestone and for a similar porosity the minimum residence time to generate the ^4He in this formation is 60,000 yr. The generation of the observed ^4He content of the thermal water, within the Carboniferous Limestone, it is impossible for a residence time as short as 10,000 yr. At least a proportion of the thermal water must be much older and may be derived from the Old Red Sandstone.

Hydraulic controls and local conditions

The thermal water probably rises at Bath by way of a southerly-dipping east-west thrust fault in the vicinity. During the artesian flow of thermal water in Kingsmead Street¹⁷, it was noted that an immediate and considerable decrease in flow occurred at the King's Spring and at the Cross Bath and Old Royal Springs. This indicates that the storage capacity of the Trias cannot be large. It was found that groundwaters which are chemically similar to Bath thermal water occur in the Trias at up to 2 km from the King's Spring. It therefore appears that the rising thermal water, under pressure, is recharging the Triassic sequence, probably by way of the Palaeozoic unconformity. This suggestion is supported by the available piezometric data (Fig. 1).

Limits on the likely hydraulic gradient between the Mendips and the Bath Spring are set by the level of the perennial

discharge springs to the north of the pericline and the maximum level of the water table on Mendip. For a first approximation, a hydraulic gradient of 0.005 appears likely. Applying Darcy's law to the flow rate at Bath for an estimated cross-sectional area perpendicular to the direction of flow of $7 \times 10^5 \text{ m}^2$ (the Carboniferous Limestone being ~700 m thick over a 1 km segment), the effective permeability controlling the flow of thermal water at depth is calculated as $\sim 0.3 \text{ m day}^{-1}$. This is reasonably lower than the permeability to diffuse flow of 4 m day^{-1} in the Mendip Carboniferous Limestone outcrop²³. For an effective porosity between 1 and 10%, volumetric considerations suggest the actual flow velocity would be between 0.1 and 0.01 m per day. Up to 4,000 yr would therefore be required to traverse the 15 km between Mendip and Bath.

Discussion

The thermal water at Bath has very constant chemical and thermal characteristics. It is of meteoric origin, and was recharged in climatic conditions similar to those in the Holocene. Recharge probably takes place on the Mendips, where sufficient elevation provides the piezometric head controlling the flow across the basin. The Carboniferous Limestone acts as the last important storage aquifer for the thermal water where it equilibrates chemically and thermally at a depth of between 2,700 and 4,300 m at a temperature in the range of 64–96 °C, before relatively rapid ascent to the spring discharge points. At Bath, the thin Triassic aquifer which overlies the Palaeozoic unconformity is recharged by the rising thermal water but the major proportion discharges at the King's Spring through the Lower Lias. Although various techniques have been used to try to determine its residence time no unambiguous estimate can presently be made.

Although the hydraulic model is simplistic, it is in order of magnitude agreement with the conclusion that the bulk of the thermal water is <10,000 yr old and on balance we favour this interpretation. To account for the observed level of ^4He , this implies that a small proportion of the thermal water must have an extremely long residence history, probably in the Old Red Sandstone, before chemical re-equilibration in the Carboniferous Limestone.

Implications for geothermal development

At the King's Spring, Bath, a discharge of almost $50 \text{ m}^3 \text{ h}^{-1}$ of water at 46 °C results from a natural head loss of 10 m. A pumped drawdown of 30 m in a borehole tapping the resource at depth might therefore provide a supply of $100\text{--}110 \text{ m}^3 \text{ h}^{-1}$. If the actual discharge were at 70 °C and the water were rejected from a heat exchanger at 25 °C, this would give a thermal yield of about 5 MW. The use of a heat pump would increase the thermal yield.

It is probable that the Carboniferous Limestone stores thermal water elsewhere along its 15-km strike between Bath and

Bristol. This is, however, only speculation and could only be proved by exploratory drilling.

It may be possible to exploit thermal groundwater stored in the Triassic aquifer at Bath without the need for wildcat drilling in the Carboniferous Limestone. The low effective bulk permeability of the Carboniferous Limestone at depth, however, is unlikely to support a flow of thermal water (across a 1 km wide section) much greater than the natural flow at the Bath springs. There could, therefore, be no exploitation of thermal water within a radius of several kilometres from Bath without risk of seriously affecting the spring discharge. The hydrogeological conditions at Bath must be thoroughly investigated before any such development, if the intrinsic historical and tourist value of the springs is to be protected.

Received 19 January; accepted 19 May 1982.

1. *Final Rep. to EEC under contract 084-76-EG(UK)* (Institute of Geological Sciences, London, 1980).
2. Welch, F. B. A. & Crookall, R. *British Regional Geology: Bristol and Gloucester District* (HMSO, London 1948).
3. Dewar, J. & Strutt, R. J. *Reports to the Corporation of Bath* (1904).
4. Munro, J. M. H. *Report to the Corporation of Bath*. (1928).
5. Burgess, W. G., Edmunds, W. M., Andrews, J. N., Kay, R. L. F. & Lee D. J. *Inst. geol. Sci. Rep. to EEC under contract 084-76-EG(UK)* (1979).
6. Riley, J. P. J. *Chem.* **11**, 190-192 (1961).
7. Judd-Lewis, S. in *Bath Official Handbook* (Mendip, Bath, 1938).
8. Heriot, M. *Proc. Bath Field Club* **3**, 171-177 (1877).
9. Ekin, C. in *The Thermal Baths of Bath* (ed Freeman, H. W.) (Hamilton, Adams and Co, London, 1888).
10. Plummer, L. N., Jones, B. F. & Truesdell, A. H. *U.S. geol. Surv. Wat. Res. Inv.* 76-13 (1976).

We thank the City Councils of Bath and Bristol for permission to carry out work at the springs and for their assistance during sampling, the Wessex Water Authority for their collaboration in the regional groundwater sampling, and local landowners for providing access. We particularly thank colleagues at the IGS for analytical work and considerable help both during the sampling and in discussion of the results. We also acknowledge the help of SURRC, East Kilbride and AERE, Harwell for providing the radiocarbon and tritium analyses. We thank Dr G. A. Kellaway for valuable discussions concerning the geology of the region. The work was funded partly by the U.K. Department of Energy and the EEC under research contract 084-76-EG(UK) and partly by NERC research grant GR3/2622A. It is published with permission of the Director, Institute of Geological Sciences (NERC).

11. Osmond, J. K. & Cowart, J. B. *Atom. Energy Ref.* **14**, 621-679 (1976).
12. Andrews, J. N. & Wood, D. F. *Inst. Min. Metall. Trans.* **B81**, B197-210 (1972).
13. Andrews, J. N. & Wood, D. F. *Phil. Mag.* **27**, 307-310 (1974).
14. Masson, I. & Ramsey, Sir W. *Trans. J. chem. Soc.* **101**, 1370-1376 (1912).
15. Craig, H. *Science* **33**, 1702-1703 (1961).
16. Bath, A. H., Edmunds, W. M. & Andrews, J. N. in *Isotope Hydrology* Vol. 2, 545-568 (IAEA Vienna, 1979).
17. Armstrong, W. *An Account of the Tapping and Closing of a Hot Spring, Bath* (Privately published pamphlet, Bristol, 1838).
18. Fourner, R. O. & Truesdell, A. H. *Geochim. cosmochim. Acta* **37**, 1255-1275 (1973).
19. Fourner, R. O. in *Proc. Symp. Hydrogeochemistry Biogeochemistry* Vol. 1, 122-139, (Clarke, Washington, 1973).
20. Kellaway, G. A. *Bull. geol. Surv. Gt Br.* No. 27, 49-153 (1967).
21. Evans, G. V., Otlet, R. L., Downing, R. A., Monkhouse, R. A. & Rae, G. in *Isotope Hydrology* Vol. 2, 679-708 (IAEA, Vienna, 1979).
22. Andrews, J. N. & Lee, D. J. *J. Hydrol.* **41**, 233-252 (1979).
23. Atkinson, T. C. *J. Hydrol.* **35**, 93-110 (1977).

T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes

Eugenio Santos, Steven R. Tronick, Stuart A. Aaronson, Simonetta Pulciani & Mariano Barbacid

Laboratory of Cellular and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

A transforming gene isolated from T24 human bladder carcinoma cells is closely related to the BALB murine sarcoma virus (MSV) onc gene (v-bas). This transforming gene is localized to a 4.6 kilobase pair (kbp) region and is expressed as a 1.2-kbp polyadenylated transcript, which contains v-bas related sequences. Moreover, antisera known to detect the immunologically related onc gene products of BALB- and Harvey-MSVs recognized elevated levels of a related protein in T24 cells. The normal human homologue of v-bas was found to be indistinguishable from the T24 oncogene by heteroduplex and restriction enzyme analysis. These results imply that rather subtle genetic alterations have led to the activation of the normal human homologue of v-bas as a human transforming gene.

THE genetic alterations that cause normal cells to become malignant in natural conditions are for the most part still obscure. A potentially important approach towards understanding such mechanisms has come from studies of acute transforming retroviruses. It is now well established that the majority of these viruses possess substituted viral genes necessary for replication with a discrete segment of host genetic information¹. When incorporated within the retroviral genome, such transduced cellular (*onc*) sequences acquire the ability to induce neoplastic transformation. At least 24 independent acute transforming retrovirus isolates have been found to date, of which some contain the same or closely related *onc* genes²⁻⁷. These findings have suggested that vertebrates contain a limited number of cellular genes that can acquire transforming properties when recombined with retroviral sequences.

The development of DNA-mediated gene transfer techniques has recently made it possible to identify transforming genes in a variety of tumours, including some of human origin⁸⁻¹⁹. DNAs

isolated from either naturally occurring human tumours or from human tumour cell lines, have been shown to induce morphological transformation when transfected into suitable assay cells¹²⁻¹⁹. In fact, a transforming gene has been isolated by molecular cloning techniques from the T24²⁰ and EJ²¹ human bladder carcinoma cell lines^{19,22,23}. Preliminary characterization of this bladder carcinoma oncogene has revealed it to be of human origin and small size (less than 6 kilobase pairs (kbp))^{19,22,23}. We undertook the present studies to investigate whether this human bladder carcinoma oncogene was related to any of the cell-derived *onc* sequences that have acquired neoplastic properties on transduction into retroviruses.

A human bladder carcinoma transforming gene is related to v-bas

The availability of a molecularly cloned bladder carcinoma transforming gene, designated the T24 oncogene, enabled us to examine its relationship with the *onc* genes of a variety of

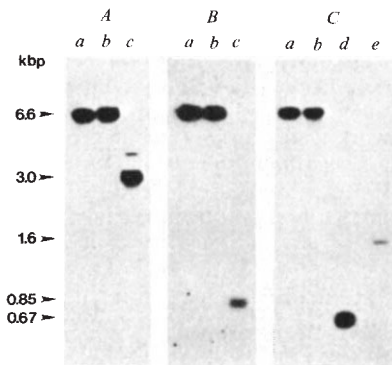


Fig. 1 Relationship between the T24 oncogene and the retroviral *onc* genes of BALB-MSV (*v-bas*) and of MC29 virus (*v-myc*). Molecularly cloned DNAs, including 100 ng of a 6.6-kbp *Bam*HI fragment containing the T24 oncogene, purified from λ T24-15A DNA¹⁹ (a), 100 ng of a 6.6-kbp *Bam*HI fragment containing the T24 oncogene purified from pT24-C3³² (b), 200 ng of the above 6.6-kbp *Bam*HI fragment but digested with *Sac*I (c), 100 ng of 0.67-kbp *Hind*III to *Bam*HI fragment of *v-bas* (d), and 200 ng of 1.6-kbp *Sac*I to *Bam*HI fragment of *v-myc* (e), were electrophoresed in horizontal agarose (1% w/v) gels, blotted to nitrocellulose filters and hybridized in stringent conditions (50% formamide, $5 \times$ SSC and 42 °C) for 16 h to 5×10^6 c.p.m. of the following nick-translated probes: A, ³²P-labelled 0.67-kbp *Hind*III to *Bam*HI *v-bas* DNA; B, ³²P-labelled 1.6-kbp *Sac*I to *Bam*HI *v-myc* DNA; C, ³²P-labelled 6.6-kbp *Bam*HI T24 oncogene DNA, as described previously⁴². Hybridized blots were exposed to Kodak XR-5 film at -70 °C in the presence of intensifier screens for 4 h (A), 3 days (B) and 4 h (C). Arrows indicate the migration of each of the hybridized DNA fragments.

transforming retroviruses. We used probes obtained from DNA clones of transforming retroviruses capable of inducing a wide range of malignancies and known to contain different *onc* genes. These included avian isolates such as the MC29 virus²⁴, avian myeloblastosis virus (AMV)²⁵ and Rous sarcoma virus (RSV)²⁶, as well as mammalian viruses, including Abelson murine leukaemia virus (A-MuLV)²⁷, BALB⁴ and Moloney^{28,29} strains of murine sarcoma virus (MSV), the Snyder-Theilen strain of feline sarcoma virus (ST-FeSV)³⁰ and simian sarcoma virus (SSV)³¹. Each retroviral *onc* gene probe was prepared from a pBR322 subclone containing only *onc* sequences.

As shown in Fig. 1, the cloned 6.6-kbp *Bam*HI human DNA fragment harbouring the T24 oncogene was readily hybridized with a DNA probe containing 675 bp of *v-bas*, the *onc* gene of BALB-MSV. In the reciprocal experiment, *v-bas* DNA was detected with similar intensity with a probe composed of the T24 oncogene (Fig. 1). Cleavage of the cloned 6.6-kbp *Bam*HI human DNA with the restriction endonuclease *Sac*I is known to generate three fragments of 3.0, 2.7 and 0.85 kbp³². Only the 3.0-kbp *Sac*I fragment was detected using *v-bas* as a probe (Fig. 1). These findings firmly established a high degree of relationship between the T24 human bladder carcinoma oncogene and *v-bas*.

No homology was detected between the T24 oncogene and any other *onc* gene tested, with the exception of *v-myc*, the *onc* gene of MC29 virus. However, the degree of relationship between the T24 oncogene and *v-myc* was at least 20–50-fold lower than that detected with *v-bas*, as measured by the relative signal intensity observed (Fig. 1). The possibility that the *v-myc* probe was detecting the subset of T24 oncogene sequences related to *v-bas* was excluded by findings that *v-bas* and *v-myc* showed no homology in identical experimental conditions (data not shown). In fact, it was possible to localize the *v-myc*-related sequences within the 0.85-kbp *Sac*I fragment of the T24 oncogene rather than within the 3.0-kbp *Sac*I fragment containing *v-bas*-related sequences (Fig. 1). Recent studies have localized the transforming region of the T24 oncogene within a 4.6-kbp DNA fragment that contains both the 3.0- and 0.85-kbp *Sac*I fragments³². Therefore, these two sets of T24

oncogene sequences, those highly related to *v-bas* and those weakly related to *v-myc*, mapped within the region required for malignant transformation.

***c-bas* (human) is a nontransforming allele of T24 human bladder carcinoma oncogene**

As normal human DNA is known to contain sequences related to a variety of retroviruses including *v-bas* and *v-myc*³³, we reasoned that characterization of normal human DNA sequences homologous to each of these transforming genes might shed light on the biological significance of their related sequences. Normal human DNA was digested with *Bam*HI and submitted to Southern blot analysis for detection of sequences related to the T24 oncogene as well as to *v-bas* and *v-myc*. The T24 oncogene and *v-bas* probes both hybridized with a 6.6-kbp *Bam*HI human DNA fragment, suggesting that the normal human *v-bas*-related sequences might represent a normal allele of the T24 oncogene (Fig. 2). In addition, *v-bas* recognized a 3.0-kbp *Bam*HI fragment known to contain another member of the *c-bas* family of genes (S.T. *et al.*, unpublished observations). We also detected hybridization of the T24 oncogene to this 3.0-kbp *Bam*HI human DNA fragment after prolonged exposure (data not shown). In contrast, the *v-myc* probe hybridized with a different subset of human DNA sequences (Fig. 2). These results established that the *v-myc*-related 0.85-kbp *Sac*I fragment of the T24 oncogene did not correspond to the human *c-myc* locus.

To compare the T24 oncogene with normal human sequences related to *v-bas* [designated *c-bas* (human)], we isolated *c-bas* (human) from a library of normal human fetal liver DNA³⁴. A recombinant λ Charon 4A phage containing a 19-kbp *Eco*RI insert of human DNA exhibited a 6.4-kbp internal *Bam*HI segment that specifically hybridized with *v-bas*. This 6.4-kbp *Bam*HI fragment was subsequently subcloned in pBR322 and a representative plasmid, designated p344, used for restriction enzyme analysis. As shown in Fig. 3, the restriction map of this 6.4-kbp *Bam*HI fragment of normal human DNA closely

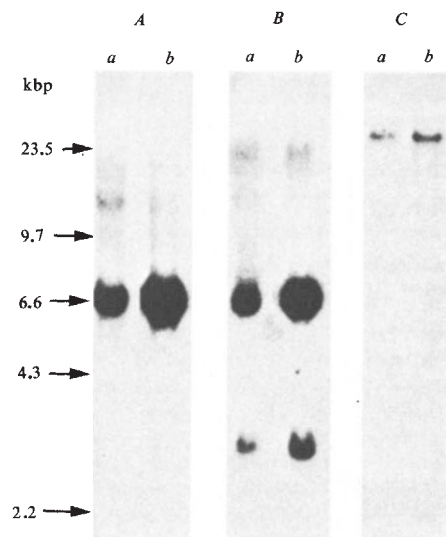


Fig. 2 Detection in human DNA of sequences related to T24 oncogene, *v-bas* and *v-myc*. Twenty microgrammes of high molecular weight DNA isolated from T24 bladder carcinoma (a) and M413 normal embryonic human cells (b) were digested with 20 U of *Bam*HI for 1 h at 37 °C, electrophoresed in horizontal agarose (0.6% w/v) gels (18 h at 30 V) and blotted to nitrocellulose filters as described previously⁴². Filters were then hybridized in stringent conditions (see Fig. 1 legend) for 44 h to 2×10^7 c.p.m. of the following nick-translated probes: A, ³²P-labelled 6.6-kbp *Bam*HI T24 oncogene; B, ³²P-labelled 0.67-kbp *Hind*III to *Bam*HI *v-bas*; C, ³²P-labelled 1.6-kbp *Sac*I to *Bam*HI *v-myc*. Hybridized blots were exposed to Kodak XR-5 film at -70 °C in the presence of intensifier screens for 2 days. Co-electrophoresed DNA fragments of *Hind*III-digested λ c1857 DNA served as size standards (labelled in kilobase pairs).

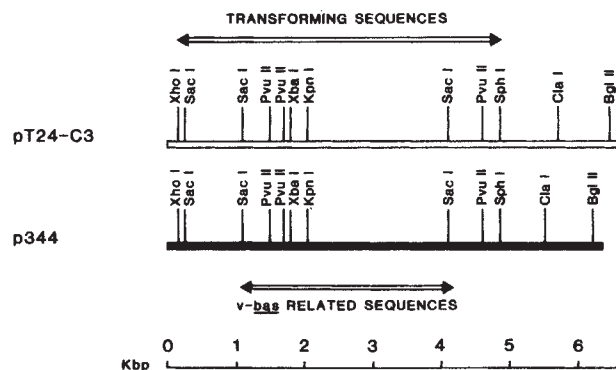


Fig. 3 Comparative restriction maps of the T24 oncogene and *c-bas* (human). The diagram depicts the location of the cleavage sites for *Xho*I, *Sac*I, *Pvu*II, *Xba*I, *Kpn*I, *Sph*I, *Cla*I and *Bgl*II restriction endonucleases within the subcloned *Bam*HI human DNA fragments containing the T24 oncogene (pT24-C3) and *c-bas* (human) (p344). The location of the transforming sequences of the T24 oncogene between the *Xho*I and *Sph*I cleavage sites has been determined elsewhere³². The location of the *v-bas*-related sequences within the 3.0-kbp *Sac*I internal fragment of both genes was determined in experiments depicted in Fig. 1.

matched that of the T24 oncogene. In fact, the only detectable difference between the two molecules was a 200-bp deletion that mapped between the *Sph*I and *Cla*I cleavage sites in *c-bas* (human) (Fig. 3). Note that this deletion maps outside the transforming sequences of the T24 oncogene³², as well as outside the sequences related to *v-bas* (Fig. 3). These results established that *c-bas* (human) is an allele of the T24 bladder carcinoma oncogene.

The T24 oncogene has been shown to transform NIH/3T3 cells efficiently, with a specific activity of $\sim 5 \times 10^4$ focus-forming units μmol^{-1} (refs 19, 22, 23). It was of obvious interest to determine whether molecularly cloned *c-bas* (human) sequen-

ces exhibited similar biological activity. As much as 1 μg of *c-bas* (human) DNA demonstrated no detectable focus-forming activity whereas in the same experiment, NIH/3T3 cells were readily transformed with as little as 1 ng of the T24 oncogene. The above results, taken together, strongly imply that the acquisition of transforming activity by the T24 oncogene must be the result of subtle genetic alterations.

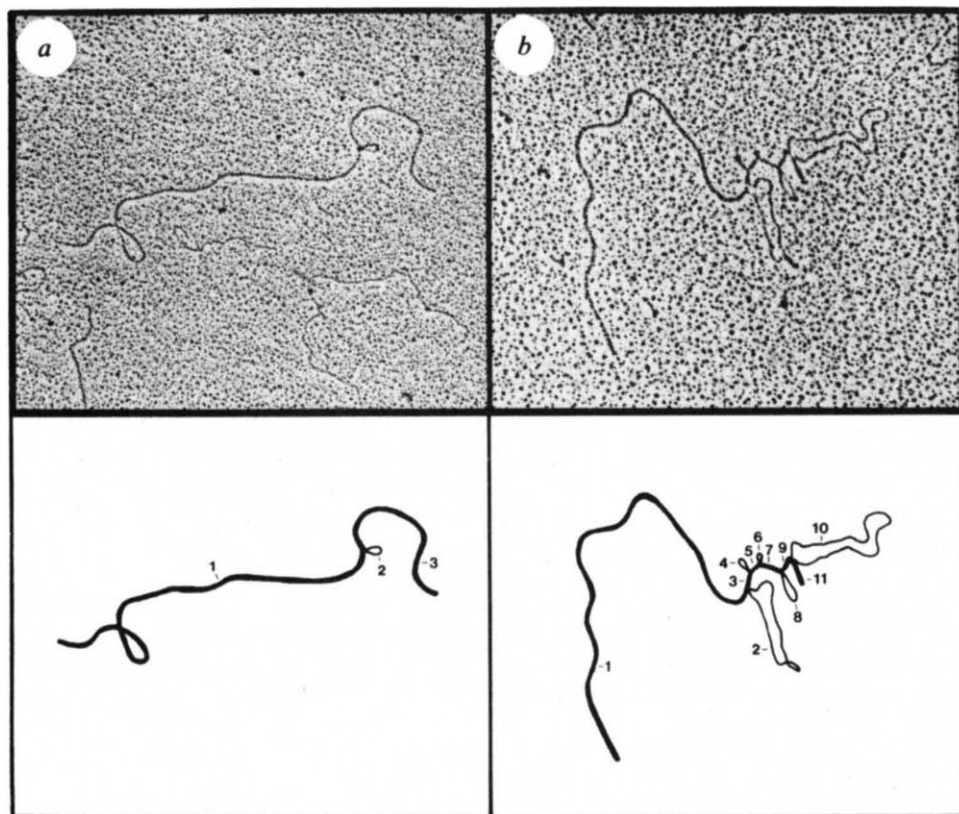
Organization of *v-bas*-related sequences within the T24 oncogene

The structural relationship between the T24 oncogene and *c-bas* (human) and *v-bas* sequences was studied by heteroduplex analysis. As shown in Fig. 4a, *c-bas* (human) and the T24 oncogene were homologous along their entire length, with the exception of a 0.2–0.3 kbp 'loop' located ~ 1 kbp from one end. These results are consistent with the restriction enzyme maps of these two molecules depicted in Fig. 3. In contrast, heteroduplexes formed between the T24 oncogene and a 0.67-kbp subclone of *v-bas* revealed the presence of a region of homology of about 0.6 kbp in length interrupted by three single-stranded regions of 0.65, 0.14 and 0.27 kbp (Fig. 4b). Its *v-bas* homologous sequences were localized within the 3.0-kbp *Sac*I fragment of the T24 oncogene, in agreement with the results obtained by Southern blot analysis (Fig. 3). The coding sequences of *v-bas* are known to reside within the viral RNA strand of the DNA (P. R. Andersen and E. P. Reddy, unpublished observations). By heteroduplex analysis (Fig. 4b) it was thus possible to orientate the T24 oncogene as left to right with respect to *v-bas* coding sequences, as shown in Fig. 3.

T24 oncogene products are related to those of the BALB- and Harvey-MSV *onc* genes

Recently, Wigler *et al.* have reported that T24 bladder carcinoma cells exhibit elevated levels of a 1.1-kbp poly(A)-

Fig. 4 Heteroduplex analysis of T24 oncogene, *c-bas* (human) and *v-bas*. Heteroduplexes were prepared as described previously⁴³. Briefly, DNAs ($2 \mu\text{g ml}^{-1}$) were denatured with alkali, annealed in the presence of 50% formamide at room temperature for 16 h and spread on to a distilled water hypophase. When pBR322 subclones were used, the supercoiled DNAs were linearized by cleavage with *Sal*I, an enzyme which does not cut any of the inserts. **a**, Heteroduplex formed between T24 oncogene and *c-bas* (human). *Bam*HI inserts purified from pT24-C3 and p344 DNAs, respectively, were used. A representative heteroduplex and an interpretative sketch are represented. Contour lengths (in kbp) of features are as follows: 1 = 5.2 ± 0.25 , 2 = 0.25 ± 0.04 and 3 = 1.15 ± 0.1 . Fragments 1, 2 and 3 correspond to the 6.6-kbp *Bam*HI DNA fragment containing the T24 oncogene. Fragments 1 and 3 correspond to the 6.4-kbp *Bam*HI DNA fragment containing *c-bas* (human). **b**, Heteroduplex analysis of T24 oncogene and *v-bas*. A representative heteroduplex between pT24-C3 and pHB1 DNA, a plasmid that contains the 0.67-kbp *Hind*III to *Bam*HI fragment of *v-bas*, and an interpretative sketch are represented. Contour lengths (kbp) of relevant features are as follows: 1 = 3.9 ± 0.2 , 2 = 1.8 ± 0.2 , 3 = 0.10 ± 0.02 , 4 = 0.27 ± 0.03 , 5 = 0.10 ± 0.02 , 6 = 0.14 ± 0.02 , 7 = 0.2 ± 0.02 , 8 = 0.65 ± 0.05 , 9 = 0.25 ± 0.03 , 10 = 3.0 ± 0.2 , 11 = 0.25 ± 0.03 . Fragments 1 and 11 correspond to pBR322 sequences. Fragments 3, 5, 7 and 9 correspond to *v-bas* sequences. Fragments 2, 4, 6, 8 and 10 correspond to the 6.6-kbp *Bam*HI DNA fragment containing the T24 oncogene.



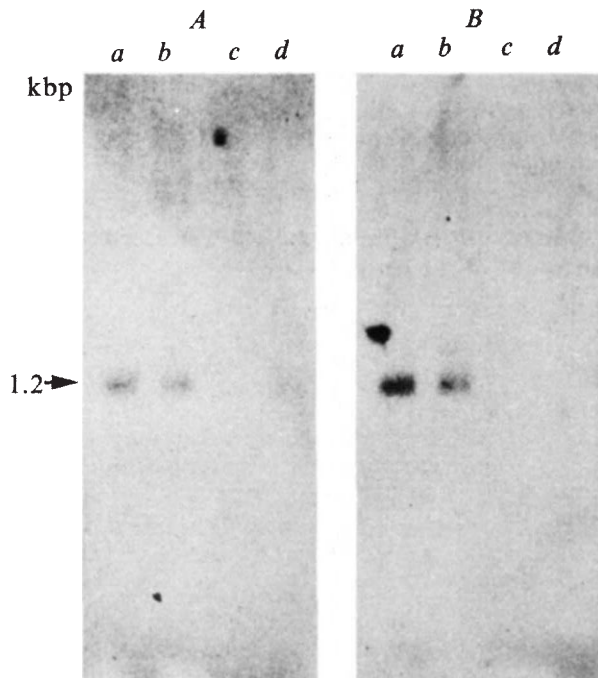


Fig. 5 T24 oncogene and *v-bas*-related transcripts in human cell lines. *a*, *b*, Poly(A)-containing RNA, isolated from T24 bladder carcinoma cells (two independent preparations); *c*, M413 normal embryo cells; *d*, HT-1080 fibrosarcoma cells were obtained as previously described^{44,45}. RNA samples (5 µg) were denatured in 13 mM methylmercury hydroxide and size fractionated in horizontal agarose (1% w/v) gels containing 5 mM methyl mercury⁴⁶. Electrophoresed gels were blotted to nitrocellulose filters which were subsequently hybridized with 10⁷ c.p.m. of the following nick-translated probes: *A*, ³²P-labelled 6.6-kbp *Bam*HI T24 oncogene; *B*, ³²P-labelled 0.67-kbp *Hind*III to *Bam*HI *v-bas*. Hybridization was performed in 50% formamide, 5 × SSC at 37 °C for 44 h. Hybridized blots were exposed to Kodak XR-5 films at -70 °C in the presence of intensifier screens for 4 days. The arrow indicates the size of the detected transcript as deduced from its relative migration with respect to ribosomal RNA size standards.

containing RNA that specifically hybridized with the T24 oncogene²⁰. This transcript was also detected in normal cells, although at much lower levels. As shown in Fig. 5A, similar results were obtained in our laboratory. To establish whether the *v-bas*-related sequences of the T24 oncogene were transcribed, poly(A)-containing RNA of T24 cells was submitted to Northern blot analysis. As shown in Fig. 5B, a 1.2-kbp RNA

molecule hybridized with both the T24 oncogene and *v-bas* probes. These results indicated that *v-bas*-related sequences are an integral part of the T24 oncogene transcript.

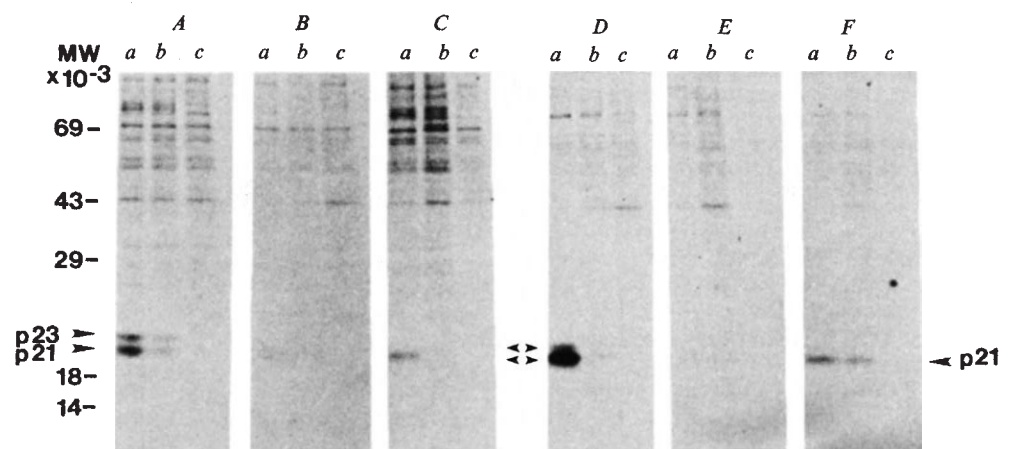
The transforming gene product of *v-bas* has been identified as a polypeptide with an apparent molecular weight of 21,000, designated p21^{bas} (ref. 4). An antigenically related protein, p21^{ras}, has been identified as the gene product of Harvey-MSV³⁵, another murine sarcoma virus whose *onc* gene is closely related to that of BALB-MSV⁴. Antisera obtained from BALB- and Harvey-MSV tumour-bearing rats were used to search for the putative T24 oncogene translational product in a variety of human tumour cells including parental T24 bladder carcinoma cells. As shown in Fig. 6, low levels of a protein migrating with an apparent molecular weight of 21,000 were detected in M413 normal human embryo fibroblasts³⁶ as well as in HT-1080 human fibrosarcoma cells³⁷. Similar results were observed with a variety of human tumour cell lines established from tumours other than bladder carcinomas (data not shown). This p21 protein was also detected in T24 cells but at somewhat elevated levels. In addition, T24 cells contained an antigenically related 23,000 molecular weight protein which was not detected in any of the other human cells analysed (Fig. 6).

To investigate further whether this protein was coded for by the T24 oncogene, we analysed NIH/3T3 cells transformed by this gene (Fig. 6D). Such transformants expressed levels of the p23 protein comparable with those detected in T24 cells. In addition, T24 oncogene-transformed NIH/3T3 cells expressed high levels of a p21 protein that co-migrated with the gene product of BALB-MSV (Fig. 6). Whether this protein represents a post-translationally modified human p23 molecule or the endogenous mouse p21 protein, remains to be determined. In either case, our findings support the concept that the T24 oncogene encodes a protein immunologically related to the transforming gene products of BALB- and Harvey-MSV.

Implications

Recent intensive investigation of acute transforming retroviruses has led to considerable knowledge concerning their genomes and translational products, as well as their interactions with the host¹. Thus, our demonstration that the T24 oncogene represents an activated human cellular analogue of a retroviral *onc* gene should make it possible to apply this knowledge to understanding the mechanisms by which the T24 oncogene induces transformation. By use of high titred antisera directed against the BALB and Harvey MSV-coded p21 proteins^{4,35}, we were able to detect an antigenically related p23 protein specifically expressed in T24 cells. It should be possible to

Fig. 6 Detection of proteins antigenically related to the *onc* gene products of BALB- and Harvey-MSV in human tumour cell lines and in NIH/3T3 cells transformed by the T24 oncogene. Exponentially growing cells (~2 × 10⁶ cells per immunoprecipitation reaction) including: *A*, T24 human bladder carcinoma cells; *B*, M413 normal human embryonic fibroblasts; *C*, HT-1080 human fibrosarcoma cells; *D*, 44-91 NIH/3T3 cells transformed by the T24 oncogene; *E*, NIH/3T3 mouse cells; *F*, NIH/3T3 cells nonproductively transformed by BALB-MSV, were labelled with ³⁵S-methionine as previously described⁴⁷. Radiolabelled cell extracts were incubated for 1 h at 4 °C with 0.005 ml of sera obtained from: *a*, Harvey-MSV-induced tumour-bearing rats; *b*, BALB-MSV-induced tumour-bearing rats; *c*, uninfected normal rats. Immunoprecipitates, collected with the aid of *Staphylococcus aureus* protein A bound to Sepharose 4B beads, were disrupted in the presence of 1% SDS and 2% 2-mercaptoethanol and analysed by electrophoresis in 8–16% linear gradient SDS-polyacrylamide gels. Electrophoresed gels were dried, fluorographed⁴⁸ and exposed to Kodak XR-5 films for 4 days at -70 °C. Molecular weight (MW) markers co-electrophoresed with the above samples included bovine serum albumin (69,000), ovalbumin, indicated by arrows.



fluorographed⁴⁸ and exposed to Kodak XR-5 films for 4 days at -70 °C. Molecular weight (MW) markers co-electrophoresed with the above samples included bovine serum albumin (69,000), ovalbumin, indicated by arrows.

establish whether this protein is present in other human bladder carcinoma cells or in other naturally occurring tumours. Such an approach may help in determining whether the T24 oncogene has an aetiological role in human neoplasia.

Detection of dominant human transforming genes by DNA-mediated gene transfer techniques has generally required the use of the NIH/3T3 cell line as the recipient assay cell^{12-16,18,19}. As NIH/3T3 is a continuous, heteroploid cell line³⁸, it is possible that genes which transform this cell line might be unable to transform normal cells. It is known that both *v-bas* and *v-ras* are capable of inducing a variety of malignancies in animals³⁹⁻⁴¹. Thus, it seems likely that when the human cellular homologue of these *onc* genes is activated, as in the case of the T24 oncogene, this gene would possess the capacity to transform normal cells as well.

The number of different genes that are activated as dominant transforming genes in human tumours is not known¹²⁻¹⁹. DNAs isolated from carcinomas of the colon and pancreas as well as from a fibrosarcoma and a rhabdomyosarcoma have been shown to transform NIH/3T3 cells in transfection assays (ref. 19 and M.B. *et al.*, unpublished). Although none of these putative human oncogenes has been found to correspond to the human homologues of retroviral *onc* genes used in the present studies,

our findings that *c-bas* (human) is a nontransforming allele of the T24 oncogene suggest that additional relationships between human oncogenes and retroviral *onc* sequences may be uncovered.

It has been well established that the cellular sequences transduced by retroviruses require viral genetic information in order to induce malignancy. The present studies demonstrate that at least one of these cellular sequences, *c-bas* (human), can be activated by a mechanism not requiring recombination with retroviral sequences. Detailed restriction enzyme analysis of this gene and its transforming allele, the T24 oncogene, has not revealed detectable differences. Thus, we must conclude that subtle genetic changes have caused the activation of this gene. Comparative sequence analysis of the T24 oncogene and *c-bas* (human) will ultimately define, not only the nature, but the number of genetic changes that led to the acquisition of transforming activity by *c-bas* (human) in T24 bladder carcinoma cells.

We thank Anne V. Lauver, Carole K. Lengyel and Linda K. Long for technical assistance, Holly Fuhrer for photographic work, and the following investigators for providing *onc* gene probes: T. Papas (*v-myc*), M. Baluda (*v-myb*) and G. Cooper (*v-src*).

Received 7 May; accepted 8 June 1982.

1. Cold Spring Harb. Symp. quant. Biol. **44**, pt 2 (1980).
2. Frankel, A. E., Gilbert, J. H., Porzig, K. J., Scolnick, E. M. & Aaronson, S. A. *J. Virol.* **30**, 821-827 (1979).
3. Shibuya, M., Hanafusa, T., Hanafusa, H. & Stephenson, J. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6536-6540 (1980).
4. Andersen, P. R. *et al. Cell* **26**, 129-134 (1981).
5. Barbacid, M., Breitman, M. L., Lauver, A. L., Long, L. K. & Vogt, P. K. *Virology* **110**, 411-419 (1981).
6. Beemon, K. *Cell* **24**, 145-153 (1981).
7. Wang, L. H. *et al. J. Virol.* **40**, 258-267 (1981).
8. Shih, C., Shilo, B. Z., Goldfarb, M. P., Dannenberg, A. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5714-5718 (1979).
9. Cooper, G. M., Okenquist, S. & Silverman, L. *Nature* **284**, 418-421 (1980).
10. Cooper, G. M. & Neiman, P. E. *Nature* **287**, 656-659 (1980).
11. Hopkins, N., Besmer, P., DeLeo, A. B. & Law, L. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7555-7559 (1981).
12. Lane, M. A., Sainten, A. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5185-5189 (1981).
13. Shih, C., Padhy, C., Murray, M. & Weinberg, R. A. *Nature* **290**, 261-264 (1981).
14. Krontikis, T. G. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1181-1184 (1981).
15. Murray, M. J. *et al. Cell* **25**, 355-361 (1981).
16. Perucho, M. *et al. Cell* **27**, 467-476 (1981).
17. Smith, B. L., Anisowicz, A., Chodosh, L. A. & Sager, R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1964-1968 (1982).
18. Lane, M. A., Sainten, A. & Cooper, G. M. *Cell* **28**, 873-880 (1982).
19. Pulciani, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2845-2849 (1982).
20. Bubenik, J. *et al. Int. J. Cancer* **11**, 765-773 (1973).
21. Marshall, C. J., Franks, L. M. & Carbonell, A. W. *J. natn. Cancer Inst.* **58**, 1743-1747 (1977).
22. Goldfarb, M., Shimizu, K., Perucho, M. & Wigler, M. *Nature* **296**, 404-409 (1982).
23. Shih, C. & Weinberg, R. A. *Cell* **29**, 161-169 (1982).

24. Lautenberger, J. A., Schulz, R. A., Garon, C. F., Tschlis, P. N. & Papas, T. S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1518-1522 (1981).
25. Souza, L. M., Komaromy, M. C. & Baluda, M. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3004-3008 (1980).
26. Shalloway, D., Zelentz, A. D. & Cooper, G. M. *Cell* **24**, 531-541 (1981).
27. Srinivasan, A., Reddy, E. P. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2077-2081 (1981).
28. Vande Woude, G. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 4464-4468 (1979).
29. Tronick, S. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 6314-6318 (1979).
30. Sherr, C. J., Fedele, L. A., Oskarsson, M., Maizel, J. & Vande Woude, G. F. *J. Virol.* **34**, 200-212 (1980).
31. Robbins, K. C., Devare, S. G. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2918-2922 (1981).
32. Santos, E., Pulciani, S. & Barbacid, M. *Fedn Proc.* (in the press).
33. Eva, A. *et al., Nature* **295**, 116-119 (1982).
34. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157-1174 (1978).
35. Shih, T. Y., Weeks, M. O., Young, H. A. & Scolnick, E. M. *Virology* **96**, 64-79 (1979).
36. Aaronson, S. A. & Todaro, G. J. *Virology* **36**, 254-261 (1968).
37. Rasheed, S., Nelson-Rees, W. A., Toth, E. M., Arnstein, P. & Gardner, M. B. *Cancer* **33**, 1027-1033 (1974).
38. Jainchill, J. L., Aaronson, S. A. & Todaro, G. J. *J. Virol.* **4**, 549-553 (1969).
39. Peters, R. L., Rabstein, L. S., Van Vleck, R., Kelloff, G. J. & Huebner, R. J. *J. natn. Cancer Inst.* **53**, 1725-1731 (1974).
40. Harvey, J. J. *Nature* **204**, 1104-1107 (1964).
41. Scher, C. D., Scolnick, E. M. & Siegler, R. *Nature* **256**, 225-228 (1975).
42. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
43. Davis, R. W., Simon, M. & Davidson, N. *Meth. Enzym.* **21**, 413-428 (1971).
44. Cox, R. A. *Meth. Enzym.* **12**, 120-129 (1967).
45. Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408-1412 (1972).
46. Bailey, J. M. & Davidson, N. *Analyst. Biochem.* **70**, 75-85 (1976).
47. Barbacid, M., Lauver, A. L. & Devare, S. J. *J. Virol.* **33**, 196-207 (1980).
48. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83-88 (1974).

Synthesis and assembly of hepatitis B virus surface antigen particles in yeast

Pablo Valenzuela*, Angelica Medina* & William J. Rutter*

* Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, and † Chiron Corporation, 4560 Horton, Emeryville, California 94608, USA

Gustav Ammerer & Benjamin D. Hall

Department of Genetics, SK-50, University of Washington, Seattle, Washington 98195, USA

The surface antigen of hepatitis B virus (HBsAg) has been synthesized in the yeast Saccharomyces cerevisiae by using an expression vector that employs the 5'-flanking region of yeast alcohol dehydrogenase I as a promoter to transcribe surface antigen coding sequences. The protein synthesized in yeast is assembled into particles having properties similar to the 22-nm particles secreted by human cells.

HEPATITIS B virus¹⁻³ is a 42-nm particle (the Dane particle) consisting of a core containing the viral genome (~3,200 base pairs (bp) of partially single-stranded DNA⁴) bound to the core

protein and its own DNA polymerase⁵⁻⁹; the virus core is surrounded by a phospholipid-containing envelope carrying the major surface antigenic determinants (HBsAg). These seem to

reside mainly in a single protein, which occurs in both a glycosylated and non-glycosylated form (molecular weight (M_r) 27,000–29,000 and 23,000–25,000, respectively, as measured on SDS gels^{10,11}). Infection with hepatitis B virus (HBV) leads not only to the production of Dane particles but also to a dramatic overproduction of 22-nm large particles and filaments (the HBsAg particle) that contain the elements of the surface envelope. These HBsAg particles are about 1,000-fold more immunogenic than the unassembled HBsAg protein¹². Because of its narrow host range (humans and chimpanzees), and as it cannot be propagated in tissue culture, investigation of the structure and mechanism of infection of HBV has been severely restricted. The advent of molecular cloning has allowed the complete virus genome to be cloned and sequenced^{5–7,13–15}, and the coding regions for the different antigens have been identified^{5–7}, raising the possibility of expressing HBV genes in alternative host systems. However, in previous attempts, high level production of HBsAg-related immunogenic material was not achieved in bacteria even in systems using powerful bacterial promoters^{14,16,17}. We therefore examined yeast as an alternative host system; this organism has complex membrane systems and the ability to secrete and glycosylate proteins by processes

resembling those observed in higher cells¹⁸. Recently, yeast has also been developed to give a system which allows the introduction, maintenance and expression of foreign genes^{19–21}. Here we report the construction of an autonomously replicating plasmid containing HBsAg-coding sequences linked to the yeast alcohol dehydrogenase I promoter. Yeast transformed with these vectors synthesize and accumulate particulate material that specifically reacts with antibodies against HBsAg.

Construction of plasmids and expression of HBsAg in yeast

The HBsAg-coding sequence used in these studies was isolated on a 835-bp *TaqI*–*HpaI* fragment^{13,16}. This fragment contains 26 bp preceding the AUG which encodes the N-terminal methionine of mature HBsAg. It lacks the DNA sequence of 489 or 522 bp that encodes a presumptive precursor which can be deduced from the viral genomic sequence⁷. The HBsAg gene fragment was joined to the yeast alcohol dehydrogenase (ADHI) promoter as described in Fig. 1 legend. The plasmid

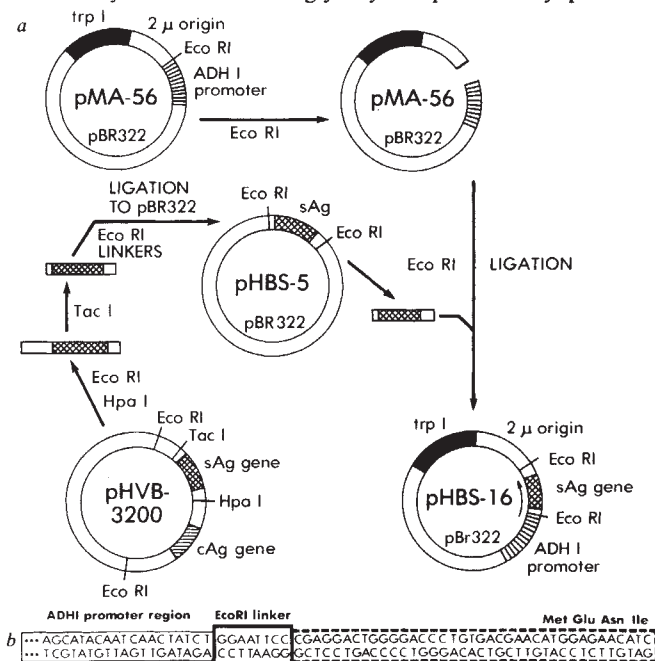


Fig. 1 **a**, Construction of plasmids for the expression of hepatitis B surface antigen in yeast. pMA-56 was constructed from YRp7 (ref. 27) by (1) removing one of its *EcoRI* sites (digestion with *EcoRI*, fill-in by DNA polymerase I Klenow fragment and blunt-end ligation); (2) replacing the yeast *ars-I* replication origin by a *PstI*–*EcoRI* fragment containing the replication origin of yeast 2 μ plasmid; and (3) replacing the smallest *EcoRI*–*BamHI* of the resulting plasmids by a 1,500-bp *EcoRI*–*BamHI* fragment containing the yeast ADHI promoter²¹. Sixty microgrammes of pHBV-3200¹³ were digested with *EcoRI* and *HpaI*, and the 1,000-bp fragment containing the surface antigen gene was isolated and digested with *TacI*; 10 μ g of the resulting 800-bp fragment were ligated to *EcoRI* linkers and cloned in the *EcoRI* site of pBR322 (pHS-5). The *EcoRI* fragment containing the HBsAg was isolated from pHS-5 by preparative gel electrophoresis after digestion with *EcoRI*, and ligated to pMA-56 (previously linearized by *EcoRI* and treated with bacterial alkaline phosphatase). After transformation of *E. coli* cells (RR1) and isolation of DNA by a small scale procedure²⁸, recombinant plasmids were analysed by cleavage with restriction endonucleases. Plasmids in which the HBsAg gene was in line with the ADHI promoter (pHS-16), and those with these regions in the opposite orientation (pHS-20) were amplified and used to transform strain XV610-8C *ade2 ade6 leu2 lys1 trp1 can1* as described elsewhere¹⁹. **b**, DNA sequence of the ADHI-HBsAg gene junction present in pHS-16. The sequence was obtained by the method of Maxam and Gilbert²⁹.

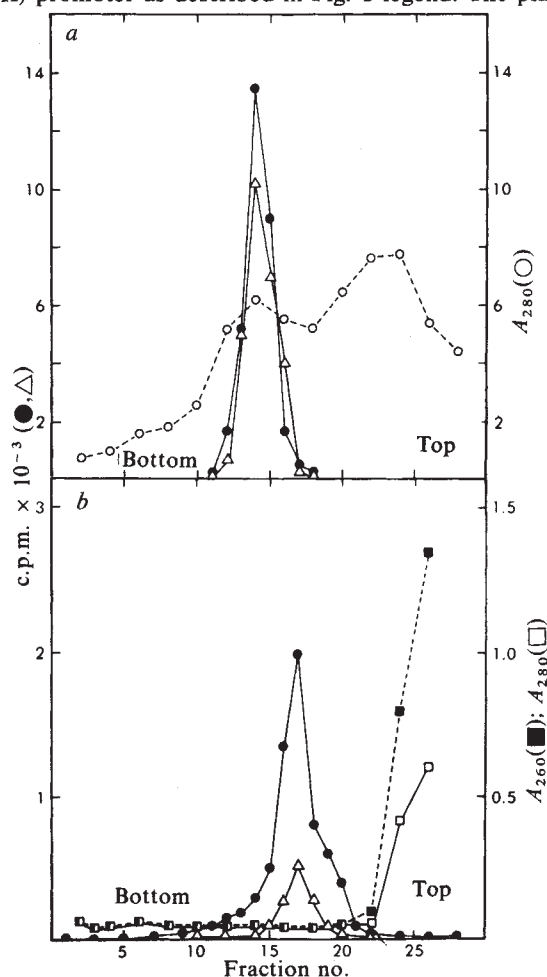


Fig. 2 **a**, CsCl gradient sedimentation of HBsAg from yeast extracts. yeast cells were grown in 0.67% yeast nitrogen base, 0.5% tryptophan-free casamino acids and 2% glucose. Cells from a 100 ml culture were pelleted, washed in 100 ml of buffer (10 mM phosphate pH 7.5, 0.1% Triton X-100), suspended in one volume of packed cells of the same buffer and broken by vortexing with glass beads (0.45–0.50 mm diameter). After centrifugation, a 0.5-ml aliquot of the extract was layered on to a 12 ml discontinuous 1.1–1.4 g cm⁻³ CsCl gradient in 10 mM phosphate buffer (pH 7.4) and run at 30,000 r.p.m. at 5 °C for 24 h in a beckman SW41 rotor. Fractions (0.5 ml) were collected from the bottom of the tube. The HBsAg activity was measured using the radioimmunoassay AUSRIA II (Abbott). The HBsAg used as standard was isolated from supernatants of PLC/PRF/5 cells^{25,26}. **b**, Sucrose gradient sedimentation of HBsAg produced in yeast cells. Active fractions from the above CsCl gradient were pooled and dialysed against 10 mM phosphate buffer pH 7.5. A 0.5-ml aliquot was layered on to a 12 ml 5–30% sucrose gradient in 10 mM phosphate buffer containing 0.15 M NaCl and run at 33,000 r.p.m. at 5 °C for 6 h in a Beckman SW41 rotor. Fractions (0.5 ml) were collected and aliquots assayed as for **a**. A 0.5-ml sample of PCL/PRF/5 cell supernatant was used as standards in the same way. ●, Yeast; △, Alexander cells for **a** and **b**.

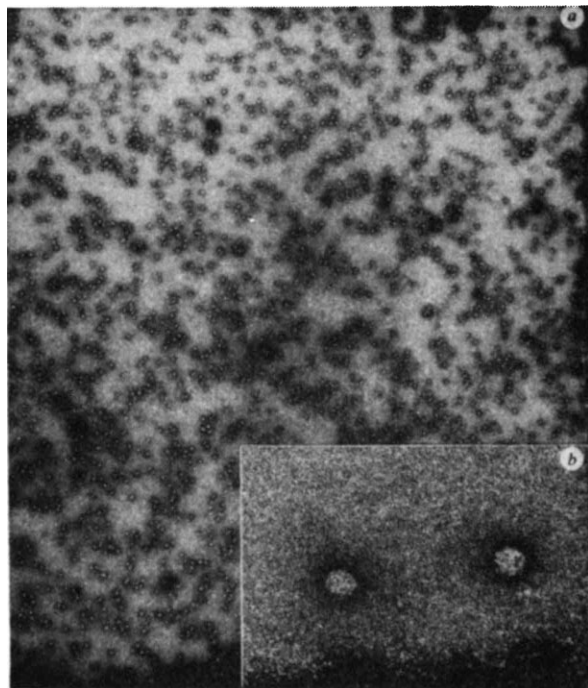


Fig. 3 Electron micrographs of HBsAg particles produced by yeast cells. *a*, Particles were purified by adsorption to a column of goat anti-HBsAg specific antibody covalently bound to Sepharose, eluted with 3 M potassium thiocyanate and visualized by negative staining with 2% phosphotungstic acid. $\times 25,000$. *b*, Higher magnification of particles purified by a combination of CsCl and sucrose gradient sedimentation and visualized after negative staining with 2% uranyl acetate. $\times 250,000$.

contains *Escherichia coli* vector pBR322, a replication origin derived from the yeast 2μ plasmid, the *trp1* gene for selection in yeast cells, and the ADHI promoter inserted into the *ter* region of pBR322. The efficacy of the ADHI promoter for expression of a linked coding sequence has recently been demonstrated by the production of interferon in this system²². Yeast cells (XV610-8C) were transformed by recombinant plasmids in which the HBsAg gene was placed either in the same 5'-3' orientation as the ADHI promoter (pHBS-16) or in the opposite orientation (pHBS-20). Cell extracts from mid-log phase cultures of these strains were assayed for HBsAg by radioimmunoassay (AUSRIA II; Abbott). A substantial level of HBsAg was detected only in the cells transformed with pHBS-16; no surface antigen was observed in cells transformed with the plasmid in which the viral gene is incorrectly oriented. The amount of surface antigen protein made per 200 ml of yeast culture was ~ 2 – $5\mu\text{g}$, calculated from the radioimmunoassay data.

Nature of viral antigen produced in yeast

The predominant form of HBsAg produced by human cells is the so-called 22-nm particle. Its biophysical properties are well documented^{23,24} and its immunological potency exceeds that of the pure protein¹². To examine the form in which surface antigen is present in yeast, we subjected extracts to equilibrium sedimentation through a discontinuous CsCl gradient. A control tube containing HBsAg from a human hepatoma cell line (Alexander or PLC/PRF/5 cells^{25,26}) was treated identically to provide a buoyant density marker. Surprisingly, HBsAg synthesized by yeast was found to band at the same density as that from the PLC/PRF/5 cell, that is, $\sim 1.2\text{ g cm}^{-3}$ (Fig. 2*a*). Peak fractions of the CsCl gradient were analysed further by velocity sedimentation in sucrose gradients. Again the peak of HBsAg produced by yeast coincided exactly with HBsAg from human cells (Fig. 2*b*), which sediments at $\sim 55\text{S}$ ²⁴.

From the sedimentation data, it is apparent that HBsAg is synthesized in yeast in the form of particles or aggregates. The nature of these particles was further characterized by electron microscopy. HBsAg synthesized in yeast was purified again by

a combination of sedimentation in CsCl and sucrose gradients. The immunoreactive material was adsorbed onto carbon film grids and negatively stained with uranyl acetate by Drs P. Bull and J. Garrido (Catholic University, Santiago, Chile). Alternatively, yeast HBsAg was purified by affinity chromatography and negatively stained with phosphotungstic acid by Drs G. Wampler, W. Ziegler, W. Miller and W. McAleer (Merck Sharp and Dohme Research Laboratories). When examined under the electron microscope, the yeast HBsAg preparation was found to contain particles (Fig. 3*a,b*) which appeared very similar to those secreted by Alexander cells. The particles purified and stained using the first method are $\sim 20\text{ nm}$ in diameter. A more extensive analysis showed that the size of the purified particles eluting with 3 M ammonium thiocyanate from an affinity column is variable, the medium size being $\sim 17\text{ nm}$ (W. Miller and W. McAleer, personal communication).

Preliminary tests have shown that these particles are antigenic, inducing antibodies in animal species (G. Buynak and W. McAleer, personal communication). It has been demonstrated previously¹² that integration of the HBsAg protein into a phospholipid structure to form the 22-nm particle dramatically increases its immunogenic properties. Thus the immunogenicity of the material synthesized in yeast is further evidence that its structure resembles that of the 22-nm particle.

The chemical composition of HBsAg molecules synthesized in yeast was determined by immunoprecipitation of ³⁵S-proteins labelled *in vivo* in a yeast strain transformed by the ADHI-HBsAg recombinant plasmid. SDS-gel electrophoresis of the anti-HBsAg immunoreactive material (Fig. 4) revealed a single band of apparent M_r 23,000, corresponding in size to the unglycosylated HBsAg component of PLC/PRF/5 cells¹¹. No higher molecular weight bands could be detected, indicating that most of the molecules were not glycosylated in these conditions.

Conclusions

Expression of hepatitis B surface antigen coding sequences in yeast leads to the production of particles immunoreactive with anti-HBsAg antibodies. These particles are similar to those

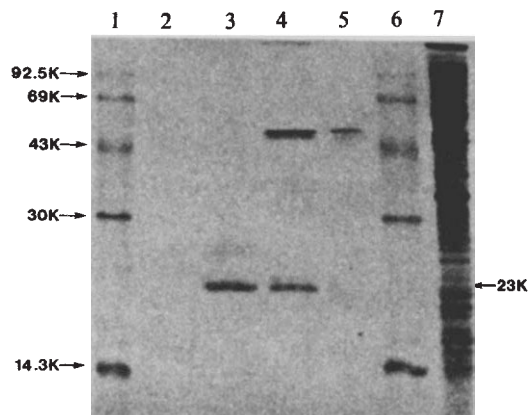


Fig. 4 Autoradiography of SDS-polyacrylamide gel of HBsAg proteins produced by yeast and PLC/PRF/5 cells. Yeast cultures and PLC/PRF/5 cells were grown in the presence of ³⁵S-methionine and ³⁵S-cysteine. Yeast cells were collected, broken by vortexing with glass beads and the resulting extracts clarified by centrifugation. Labelled HBsAg particles from PLC/PRF/5 cell supernatants were concentrated by ultracentrifugation and used as a control. Proteins from yeast and PLC/PRF/5 cells were immunoprecipitated using the SAC technique³⁰. Samples were run on 12.5% acrylamide gels according to Laemmli³¹, subjected to fluorography³², dried and exposed to X-ray film. Lanes 1 and 6, ¹⁴C-methylated proteins as molecular weight standards: phosphorylase (92,500, 92.5K), bovine serum albumin (69K), ovalbumin (46K), carbonic anhydrase (30K) and lysozyme (14.3K); lane 2, PLC/PRF/5 cell supernatants plus normal human serum; lane 3, PLC/PRF/5 cell supernatant plus human anti-HBsAg serum; lane 4, yeast extracts from strain XV610-8C with pHBS-16 plus human anti-HBsAg serum; lane 5, yeast extracts from strain XV610-8C with pHBS-16 plus normal human serum; lane 7, aliquot of total yeast ³⁵S-labelled proteins. Numbers at the left refer to protein markers as above. The number at the right indicates the 23,000- M_r protein immunoprecipitated from yeast and PLC/PRF/5 cells.

made by human carrier patients or by a hepatoma cell line; they have identical sedimentation rates and buoyant density, suggesting a similar size and ratio of protein to lipid composition for yeast and for human particles. Electron microscopic observation established that the yeast particles are more variable in size and have a smaller mean diameter than the human particles. Furthermore, unlike the HBsAg from human cells, the yeast particle does not contain significant quantities of a higher molecular weight glycoprotein, indicating that glycosylation is required neither for the formation of the particulate structures nor for immunogenicity.

The construction we have used in the HBsAg-expressing plasmid eliminates that part of the gene encoding the leader peptide of putative pre-HBsAg⁷. Therefore, formation of particles in yeast implies that the polypeptide sequence of the surface antigen molecule contains the requisite instructions for assembly of apparently normal particles within the cytoplasmic milieu of the yeast cell. The leader region and other possible precursors formed by its translation are not required for the formation of this structure. Furthermore, no other hepatitis gene products or liver-specific cell products are necessary to form the particulate structure. It seems likely, therefore, that such particles can be formed in many other heterologous cell systems, perhaps even in bacteria. It remains to be determined where within the yeast cell the 22-nm particle is assembled, and what biochemical processes accompany it. An immediate advantage of the yeast system is that the physiological and genetic requirements for particle assembly may be systematically explored, making use of yeast mutants conditionally defec-

tive for the secretory pathway¹⁸ and for other aspects of cellular metabolism.

The similarity in structure of the yeast particle to bona fide 22-nm particles and the high immunogenicity in animals emphasize the possible value of the HBsAg particle as a vaccine. The yeast HBsAg particles can in principle be produced in large quantities for this purpose. Further, the complete absence of intact HBV and/or human proteins in such preparations eliminates the possibility of secondary infections or autoimmunity problems as a result of the vaccine.

Finally, our findings illustrate the versatility of recombinant DNA methodology for the production of complex structures. Previously, recombinant DNA techniques as applied to microorganisms have only been used to produce simple proteins like insulin, growth hormone, interferon and enzymes. Our experiments suggest that more complicated structures can be formed in alternative hosts like yeast provided the genes for the requisite organizing proteins are introduced and efficiently expressed. This offers an effective approach for the analysis of organelles in higher organisms.

We thank P. Bull, J. Garrido and B. Wolanski, R. Ziegler, W. Miller and W. McAleer for the electron microscopic analysis; L. Rall for the preparation of ³⁵S-labelled HBsAg particles from PLC/PRF/5 cells; and M. Worthington and S. Bektech for the construction of pMA-56. This research was funded by grants from Merck Sharp and Dohme Research Laboratories. G.A. was supported by a grant from the Max Kade Foundation. *Note added in proof:* Since the submission of this paper, particles containing a different HBV protein, the virus core antigen, have been reported in *E. coli* carrying a recombinant plasmid³³.

Received 2 March; accepted 25 May 1982.

1. Vyas, G. N., Cohen, S. N. & Schmid, R. (eds) *Viral Hepatitis. A Contemporary Assessment of Etiology, Epidemiology, Pathogenesis and Prevention* (Franklin Institute Press, Philadelphia, 1978).
2. Zuckerman, A. J. *J. Tox. env. Hlth* **5**, 275-280 (1979).
3. Prince, A. M. *Hepatology* **1**, 73 (1981).
4. Summers, J., O'Connell, A. & Millman, I. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4597-4601 (1975).
5. Pasek, M. *et al. Nature* **282**, 575-579 (1979).
6. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. *Nature* **281**, 646-650 (1979).
7. Valenzuela, P., Quiroga, M., Zaldivar, J., Gray, P. & Rutter, W. J. in *Animal Virus Genetics* (eds Fields, B. N., Jaenisch, R. & Fox, C. F.) 55-77 (Academic, New York, 1980).
8. Robinson, W. S. & Greenman, R. C. *J. Virol.* **13**, 1231-1236 (1974).
9. Hruska, J. F. & Robinson, W. S. *J. med. Virol.* **1**, 119-124 (1977).
10. Shih, J. W. K. & Gerin, J. L. *J. Virol.* **21**, 1219-1222 (1977).
11. Peterson, D. L. *J. biol. Chem.* **256**, 6975-6983 (1981).
12. Cabral, G. A. *et al. J. gen. Virol.* **38**, 339-350 (1978).
13. Valenzuela, P. *et al. Nature* **280**, 815-819 (1979).
14. Burrell, C. J., Mackay, P., Greenway, P. J., Hofschneider, P. H. & Murray, K. *Nature* **279**, 43-47 (1979).
15. Sninsky, J. J., Siddiqui, A., Robinson, W. S. & Cohen, S. N. *Nature* **279**, 346-348 (1979).
16. Edman, J. C., Hallewell, R. A., Valenzuela, P., Goodman, H. M. & Rutter, W. J. *Nature* **291**, 503-506 (1981).

17. Mackay, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 4510-4514 (1981).
18. Novick, P., Field, C. & Schekman, R. *Cell* **21**, 205-225 (1980).
19. Beggs, G. D. *Nature* **275**, 104-109 (1978).
20. Petes, P. D. *A. Rev. Biochem.* **49**, 845-876 (1980).
21. Ammerer, G., Hitzeman, R. A., Hagie, F. E., Barta, A. & Hall, B. D. *Recombinant DNA, Proc. 3rd Cleveland Symp. Macromolecules* (ed. Walton, A. G.) 185-197 (Elsevier, Amsterdam, 1981).
22. Hitzeman, R. A. *et al. Nature* **293**, 717-722 (1981).
23. Gerin, J. L., Purcell, R. H., Hoggan, M. D., Holland, P. V. & Chanock, R. M. *J. Virol.* **4**, 763-768 (1969).
24. Gerin, J. L., Holland, P. V. & Purcell, R. H. *J. Virol.* **7**, 569-576 (1971).
25. Macnab, G. M., Alexander, J. J., Letcasas, G., Bey, E. M. & Urbanowicz, J. M. *Br. J. Cancer* **34**, 509-515 (1976).
26. Alexander, J. J., Bey, E. M., Geddes, E. W. & Letcasas, G. S. *Afr. med. J.* **50**, 2124-2128 (1976).
27. Tschumper, G. & Carbon, J. *Gene* **10**, 157-166 (1980).
28. Birnboim, H. C. & Doly, J. *J. Nucleic Acids Res.* **7**, 1513-1523 (1979).
29. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 490-565 (1980).
30. Martial, J. A., Hallewell, R. A., Baxter, J. D. & Goodman, H. M. *Science* **205**, 602-607 (1979).
31. Laemmli, U. K. *Nature* **227**, 680-685 (1977).
32. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83-88 (1974).
33. Cohen, B. J. & Richmond, J. E. *Nature* **296**, 677-678 (1982).

LETTERS TO NATURE

Evidence for an oblique magnetic solar rotator

D. O. Gough

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

The observation by Claverie *et al.*¹ of fine structure in the peaks in the power spectrum of low-degree 5-min solar oscillations has been interpreted as being a result of rotational splitting. Claverie *et al.*¹ claim that their measurements imply that an appropriately weighted average $\bar{\Omega}$ of the interior angular velocity $\Omega(r)$ of the Sun is about twice the value of Ω at the surface. At first sight their claim looks doubtful, because all $2l+1$ components of the set of modes of degree l appear in the spectrum, whereas only $l+1$ of them should be detectable. However, Isaak² has recently speculated that the additional components are produced by an intense rotating magnetic core, such as had been postulated by Dicke^{3,4} to account for the

12.2-day periodic component in the Princeton oblateness data⁵⁻⁹. Isaak² pointed out that the mean period $2\pi/\bar{\Omega} \approx 15$ days of the solar interior that was inferred from the Birmingham data¹ is consistent with a central core rotating with the 12.2-day period and an outer envelope rotating with the photosphere. Moreover, his rough estimate of a few megagauss for the r.m.s. magnetic field that is required to support his conjecture is within about a factor 10 of that required by Dicke³. Thence he concluded that the Birmingham data¹ provide the first clear empirical evidence for an intense internal solar magnetic field. Here I examine this evidence in more detail, and show that the conclusion is premature. If the magnetic core does exist, the 5-min oscillations provide no clear evidence that it is rotating rapidly. Furthermore, unless one accepts a contrived magnetic-field configuration, an explanation for the $2l+1$ components of the multiplets is still lacking.

With respect to spherical polar coordinates (r, θ, ϕ) about any axis, the linear eigenfunctions of oscillation of a spherically symmetrical star of radius R are separable in radial and angular coordinates. For example, the radial component δr of the dis-

placement eigenfunction can be written:

$$\delta r(\mathbf{r}, t) = \text{Re}[\xi(r) P_l^m(\cos \theta) e^{im\phi - i\omega t}], \quad |m| \leq l \quad (1)$$

where P_l^m is the associated Legendre function. For any given value of the degree l , there is a discrete set of modes. Each mode in the set is labelled by an integer n , called the order of the mode, which is equal to the number of zeros of $\xi(r)$ for $0 < r < R$, ($0 \leq r < R$ when $l = 0$), except possibly when n is small. The frequencies of the p modes, which are the subject of this discussion, increase with n .

The degree l measures the total horizontal wavenumber. The angular order m measures its azimuthal component about the coordinate axis. Unlike m , l is independent of the orientation of the coordinate axis; indeed, any spherical harmonic of degree l transforms under coordinate rotations into a linear combination of the harmonics of that degree. Evidently the eigenfrequencies of a spherically symmetrical star cannot depend on the orientation of the axis. Therefore the frequencies ω and eigenfunctions $\xi(r)$ are independent of m .

Unless $m = 0$, the representation (1) is of a wave propagating about the axis with angular phase speed $m^{-1}\omega$. Since ξ and ω are independent of m , it is possible to combine two such waves with equal and opposite m to form a standing wave.

By comparing with theory the power spectra of the observed low-degree oscillations^{1,10-17} it has been possible to identify the modes. In particular, the modes isolated by Claverie *et al.*¹ lie in the range: $0 \leq l \leq 2$, $16 \leq n \leq 27$. For this discussion, the values of n are sufficiently large for asymptotic theory^{18,19} to be adequate. Furthermore, it is sufficient to assume the oscillations to be adiabatic, with ξ and ω real.

Any deviation of the equilibrium state of the star from spherical symmetry is liable to destroy the simple structure of the eigenfunctions exhibited in equation (1). In the Sun it is known that symmetry is broken by rotation. It is also broken by magnetic fields. The effect on the eigenfrequencies is to split the degeneracy with respect to m . The question under investigation here is whether the splitting that has been observed in the Sun compels us to conclude that strong magnetic fields are present.

Let us consider first the effect on the modes of rotation alone. Viewed from a coordinate frame rotating with angular velocity Ω_0 about the rotation axis of the Sun, and granted that the solar angular velocity Ω measured in an inertial frame is small, the principal modifications to the motion that arise are due to Coriolis forces and advection of the wave pattern by the relative angular velocity $\Omega - \Omega_0$. For high-order p modes the Coriolis effect is small; by substituting the appropriate asymptotic form¹⁹ for the eigenfunction into equations (7), (8) of ref. 20 it can be shown that advection augments the frequency by approximately

$$\Delta\omega = mI^{-1} \int_0^R c^{-1}(\Omega - \Omega_0) dr \equiv m(\bar{\Omega} - \Omega_0) \quad (2)$$

where

$$I = \int_0^R c^{-1} dr \quad (3)$$

and c is the sound speed in the equilibrium state. Here it has been assumed that Ω is a function of r alone. Thus, with respect to an inertial spherical polar coordinate system about the rotation axis, the total frequency is

$$\omega = \omega_0 + m\Omega_0 + \Delta\omega = \omega_0 + m\bar{\Omega} \quad (4)$$

where ω_0 is the frequency of the corresponding mode of an equivalent nonrotating star¹⁷. Notice that $\bar{\Omega}$ is independent of m . Therefore the frequencies of the modes of given degree l and order n are uniformly spaced, with separation $\bar{\Omega}$. In particular, it follows from equation (4) that any standing wave pattern formed by a superposition of rotationally split modes precesses without dispersion about the rotation axis of the Sun with angular velocity $\bar{\Omega}$. In the frame of the standing wave, frequency is independent of m .

This behaviour is modified when one acknowledges that Ω is really a function of both r and θ . The splitting is no longer precisely uniform, because modes with different values of m sample the latitudinal variation of Ω differently. For example, if the angular velocity of the Sun is approximated by $\Omega(r)(1 - \alpha \cos^2 \theta)$ with α constant and positive, the high-order p -mode eigenfrequencies can be shown to be

$$\omega = \omega_0 + m \left\{ 1 - \frac{\alpha}{2} (2l+1) \frac{(l-|m|)!}{(l+|m|)!} \int_{-1}^1 [\mu P_l^m(\mu)]^2 d\mu \right\} \bar{\Omega} \\ \equiv \omega_0 + m\beta \bar{\Omega} \quad (5)$$

Thus as $|m|$ increases at fixed l the splitting increases, because the eigenfunctions become more concentrated towards the more rapidly rotating equator. The concentration is accentuated as l increases, and the mean spacing, which is $[1 - (2l+3)^{-1}\alpha]\bar{\Omega}$, approaches $\bar{\Omega}$ as $l \rightarrow \infty$. However, when l is small the variation is also small: taking $\alpha \approx 0.2$ yields a splitting between quadrupole modes with $m = 1$ and $m = 2$ only 6% greater than the splitting between those with $m = 0$ and $m = 1$, and roughly equal to the splitting of the dipole modes.

According to Claverie *et al.*¹, the major peaks in the power spectrum of their data are either single, or split into approximately uniformly spaced triplets or quintuplets. The mean spacing of the triplets is $\sim 0.67 \mu\text{Hz}$ synodic, and that of the quintuplets, which are less clearly resolved, is $\sim 25\%$ greater. The overall mean spacing is $0.75 \mu\text{Hz}$.

The signal obtained by Claverie *et al.* is a measure of the time-varying shift in a solar spectrum line. The light is integrated over the entire solar disk²⁰. Therefore, if the solar rotation axis were perpendicular to the line of sight, pure modes for which $l+m$ is odd could not contribute to the signal. This follows directly from the symmetry properties of tesseral harmonics. Thus, of all the modes of degree l , no more than $l+1$ of the $2l+1$ possible values of m should be detected. At the time of the observations the Sun's axis was tilted from the plane of the sky by about 6° . As pointed out by Isaak², this is insufficient to have produced a detectable signal from the otherwise invisible modes, unless of course, modes with odd $l+m$ systematically have 10 times the amplitude of the others.

If one were to assume with Claverie *et al.*¹ that the fine structure in the Birmingham power spectrum is rotational splitting, the most naive interpretation would be that the triplets correspond to $l = 2$ and the quintuplets to $l = 4$. Then the mean splitting between adjacent components of the multiplets would be $2\beta\bar{\Omega} \approx 2\bar{\Omega}$, implying a mean interior rotation rate of $\sim 0.38 \mu\text{Hz}$. This is similar to the rotation rate of the photosphere, which is $0.42 \mu\text{Hz}$ synodic at the equator.

This interpretation faces several difficulties, such as the problem of explaining why doublets presumed to be associated with dipole modes (to which the observations are most sensitive¹⁷) are not prominent in the spectrum. But most important is that it is contradicted by the mode identification made previously^{1,10,13-15}, which depends on the qualitative structure of the pattern apparent in the frequencies of the modes^{13-18,21}. That pattern, which repeats as n increases, consists of approximately uniformly spaced groups of modes, of alternately odd and even degree. It is a basic signature of the eigenvalues of the wave equation that describes asymptotically high-order modes in any spherically symmetrical system. Though the absolute values of the theoretical eigenfrequencies are dependent on the solar model, the pattern in which they are distributed is a basic geometrical property. It is a consequence of the fact that the Sun is essentially a sphere (and not, for example,

Table 1 Frequencies of high-order dipole p modes in a rotating magnetic star

m'	$m = -1$	$m = 0$	$m = +1$
0	$\omega_0 + \delta_0\omega - \beta\bar{\Omega}$	$\omega_0 + \delta_0\omega$	$\omega_0 + \delta_0\omega + \beta\bar{\Omega}$
1	$\omega_0 + \delta_1\omega - \beta\bar{\Omega}$	$\omega_0 + \delta_1\omega$	$\omega_0 + \delta_1\omega + \beta\bar{\Omega}$

a cube). Therefore, whilst one might suspect that n has been incorrectly assigned, the identification of l is robust: the triplets found by Claverie *et al.* have $l = 1$, and the quintuplets $l = 2$. So why is it that there are more components in the observed spectrum than should be expected?

Of course, all the components would be detectable if the solar interior were rotating about an axis that is inclined to that of the photosphere. In that case, because it would be quite unlikely that the axis of mean rotation would precess about the photospheric axis with the same angular velocity as the Earth's orbital rotation about the Sun, the relative heights of the rotationally split components in the power spectrum would be modulated on a time scale of a year. This could be checked by observation.

The presence of the central dipole component strongly suggests to Isaak² the existence of a perturbation which is inclined at a large angle from the axis of material rotation, and which precesses about it. Isaak contends that this perturbation is produced by an internal oblique magnetic rotator of a kind proposed previously by Dicke^{3,4}.

Both the magnetic effects and the rotational effects are small, producing frequency changes of $< 0.1\%$. Therefore, as Isaak² acknowledged, it is quite adequate to linearize about the nonrotating nonmagnetic state, and treat the two perturbations separately. To this end it is conceptually convenient to work in a frame S' rotating with the magnetic field, say with angular velocity Ω_0 , and transform to the inertial frame later. In S' the equilibrium state is steady, and the oscillations about that state can be described by eigenfunctions that are separable in space and time. I shall assume, with Dicke³, that the field has an axis of symmetry, in a direction (Θ, Φ) with respect to an inertial coordinate system S whose axis coincides with the axis of rotation. The axis of magnetic symmetry is taken as the axis of spherical polar coordinates (r, θ', ϕ') in S' . With respect to these coordinates the angular order of a mode will be denoted by m' .

Perturbation theory for magnetic stars has been developed by Ledoux and Simon²², and subsequently by Goossens²³. Both the equilibrium state and the normal modes of oscillation are expanded in powers of a parameter λ , which characterizes the strength of the Lorentz force in relation to the pressure gradient; it measures both the deviation from spherical symmetry of the equilibrium configuration and the perturbation to the eigenfrequencies. In the Sun λ is small: if, for example, the magnetic field is confined to the core, and has intensity 5×10^7 G, then $\lambda \approx 10^{-3}$. Since the postulated magnetic field is buried in the solar interior, singular perturbations in the surface layers^{24,25} are of no consequence.

Consider first the perturbation to the shape of the dipole wave pattern. Formally, the most striking effect of the magnetic field is to resolve the degeneracy in the eigenfunctions. The dipole oscillations are separated into three well-defined modes, however weak the field. But that does not change the number of peaks in the power spectrum. A new peak associated with the $m = 0$ mode resulting from a modification to the geometry of the wave pattern can come about only by coupling to a mode of even degree l^* . If ω^* is the frequency of free oscillation of the coupled mode that is nearest to ω , the ratio of the amplitude of the coupled mode to that of the dipole mode is of order $\sigma = \min(1, \lambda\omega/|\omega - \omega^*|)$. If $l^* = 0$ or $l^* = 2$, $|\omega - \omega^*| \approx 70 \mu\text{Hz}$. Then $\sigma \approx 40\lambda$, which is probably too small for the coupled mode to be detected. If $l^* > 2$ it is conceivable that a chance resonance could make $\sigma = 0(1)$, but in that case the sensitivity of the observations is too low to account for the observed central peak¹⁷. Similar arguments apply to the quadrupole modes. Therefore the most significant influence the magnetic field is likely to exert is to change the speeds of propagation of the waves.

At this juncture it should be pointed out that any standing dipole mode has an axis of symmetry, and can be represented as a linear combination of the three modes with $l = 1$. If somehow that pattern could be caused by the rotating magnetic field to rotate with angular velocity Ω_1 about the axis of the

inertial frame S , the sole effect would be to augment the frequencies of the three components in S by $m\Omega_1$. After adding the effect of material rotation, the frequencies of the modes would become $\omega_0 + m(\Omega_1 + \bar{\Omega})$. This, of course, is completely analogous to, and observationally indistinguishable from, rotational splitting, and therefore could not introduce another component to the power spectrum.

The effect of the magnetic field $\mathbf{B}$, however, is not to induce a simple rotation of a standing mode. Because the Lorentz force is quadratic in the field, changing the sign of $\mathbf{B}$ has no effect. Hence there is no distinction between waves propagating positively and negatively about the axis of symmetry. Though frequency changes can depend on the magnitude of m' , they cannot depend on its sign. In S' , combinations of modes with like $|m'|$ yield standing waves that do not precess; but unlike rotationally split modes, the frequency splitting is sensitive to m' , at least when l is small. It is an immediate corollary that to leading order in λ the frequencies observed are not influenced by the angular velocity of the buried magnetic field configuration: as is evident in equation (4), after adding the effect of rotation and transforming to the inertial frame (or, for that matter, to any other rotating frame that does not depend on Ω_0 , such as the frame rotating with the Earth), Ω_0 cancels.

It is not inconceivable, however, that an observable dipole triplet could be induced by a magnetic field. Let us represent any mode of degree l by a superposition of the modes in S' of degree l and angular orders m' . Then the effect of the field is to augment the frequencies by $\delta_{m'}\omega$, say, where $\delta_{-m'}\omega = \delta_{m'}\omega$. After transforming the modes to S , and adding the effect of rotation, each component (l, m) is found to have multiple frequencies. These are listed in Table 1 for the dipole modes. Since in this case the $m = 0$ modes cannot be detected, there are four frequencies that should appear in the spectrum. It is possible, however, to reduce these to a triplet, if the two inner components have identical frequencies. This occurs if

$$|\delta_1\omega - \delta_0\omega| = 2\beta\bar{\Omega} \quad (6)$$

The frequency splitting is then $2\beta\bar{\Omega}$, which is again consistent with almost uniform rotation.

The hypothesis raises several issues. For example, condition (6) is a relationship between the magnetic field and the rotation, and was engineered to force agreement with observation. If it were satisfied in the Sun, it would be unlikely to be accidental. We know, of course, that there is a dynamical connection between Ω and $\mathbf{B}$. But why should it be such that particular oscillation frequencies coincide, especially since the energy in the oscillations is very much less than the field energy? Moreover, we must acknowledge that $\delta_{m'}\omega$ might vary with frequency, and question whether equation (6) can be satisfied sufficiently well by all the relevant modes. The observations would therefore set limits on how rapid the variation could be. There are also further conditions, analogous to equation (6), to be explained. They arise from demanding that quadrupole modes are split into only quintuplets. These would constrain Ω and $\mathbf{B}$ further.

It is also of interest to compare the amplitudes of the components of the multiplets. It can be shown that if the three dipole modes in S' have r.m.s. displacement amplitudes $A_{m'}$ in the atmosphere, the mean squared amplitudes in S , in ascending or descending order of frequency according to whether $\delta_1\omega$ is

Table 2 Frequency shifts (μHz) of those high-order solar quadrupole p modes that can be detected by whole-disk observations

m'	$m = -2$	$m = 0$	$m = +2$
0	0.91	1.68	2.45
1	3.08	3.85	4.62
2	1.56	2.33	3.10

Observations are referred to a frame rotating with the Earth about the Sun and computed assuming the presence of a toroidal magnetic core chosen so as to reduce the dipole quadruplet to a triplet. Rotational splitting was computed from equation (5) with $\alpha = 0.2$.

greater or less than $\delta_0\omega$, should be in the ratios

$$\begin{aligned} A_0^2 \sin^2 \theta : \\ A_0^2 \sin^2 \theta + \frac{1}{2}(A_1^2 + A_1^2)(1 - \frac{1}{2} \sin^2 \theta) : \\ \frac{1}{2}(A_1^2 + A_1^2)(1 - \frac{1}{2} \sin^2 \theta) \end{aligned} \quad (7)$$

Because the whole-disk observations are equally sensitive to modes with $m = -1$ and $m = +1$, these represent the ratios of the power in each peak of the triplet. Note that whatever the values of A_m and θ , the power in the central peak is the sum of the powers in its two neighbours. This appears not to be in accord with observation¹.

The magnetic core postulated by Dicke^{3,4} is presumed to have been formed by field stretching produced by the differential rotation. This created an essentially toroidal field, about an axis which subsequently departed from the Sun's rotation axis, in a manner that has been discussed by Mestel *et al.*²⁶⁻²⁸. Now it lies almost in the equatorial plane, with $\theta = 86^\circ$. I approximate the magnetic field by

$$\mathbf{B} = (0, 0, B(r) \sin \theta' \cos \theta') \quad (8)$$

without questioning its stability. This is the form originally adopted by Dicke³, though a subsequent analysis of the oblateness data suggests a more complicated configuration²⁹.

The perturbations $\delta_m\omega$ generated by $\mathbf{B}$ can be computed using Goossens's analysis²³. In the limit of high-order p modes the result is

$$\begin{aligned} \delta_m\omega \sim \frac{(2l+1)(l-|m'|)!}{u(R)(l+|m'|)!} \left[\int_0^R \left\{ a \left[\frac{d}{dr} (r\xi B) \right]^2 + b m'^2 \xi^2 B^2 \right\} dr \right. \\ \left. - \frac{\omega}{4\pi} \int_0^R dr \int_0^\pi c^{-1}(c_1^2 - \varepsilon) [P_{l'}^{m'}(\cos \theta')]^2 \sin \theta' d\theta' \right] \quad (9) \end{aligned}$$

where

$$u = \int_0^r c^{-1} dr \quad (10)$$

$$\xi = r^{-1} \rho^{-1} c^{-1} u^{1/2} J_{l+3/2}(\omega u) \quad (11)$$

$$a(l, m) = \int_{-1}^1 [P_{l'}^{m'}(\mu)]^2 \mu^2 (1 - \mu^2) d\mu,$$

$$b(l, m') = \int_{-1}^1 [P_{l'}^{m'}(\mu)]^2 \mu^2 d\mu \quad (12)$$

$\rho(r)$ is the density of the equilibrium model in the absence of $\mathbf{B}$, and J_ν is the Bessel function of the first kind. The variable c_1^2 is the small θ' -dependent part of the square of the adiabatic sound speed in the equilibrium configuration, measured in units of c^2 , and $\varepsilon(\theta')$ is such that $r = (1 + \varepsilon)R$ defines the equilibrium photosphere. The first integral in equation (9) represents the direct effect of the Lorentz force on the oscillations, and the second integral results from the axisymmetrical distortion to the equilibrium state.

It can now be argued that an intense large-scale magnetic field cannot pervade the entire radiative region of the Sun. If there were such a field, all the terms in the integrals in equation (9) would be of roughly the same magnitude. According to Dicke³⁻⁹ ε is of order 3×10^{-5} , and according to Hill *et al.*²⁸ it must be even smaller. This sets an upper limit to the intensity of the magnetic field. It implies that the magnitude of $\delta_m\omega$, to within a factor of order unity, can be no larger than $3 \times 10^{-5}\omega$, which is too small for equation (6) to be satisfied.

The distortion associated with a field confined to the core is relatively small at the photosphere. This permits the field to be stronger without it being ruled out by the oblateness measurements. Can it be strong enough to produce the frequency shift required? Note first that the contribution to $\delta_m\omega$ from the distortion of the equilibrium structure is proportional to ω . Therefore, if that term dominated, it would not be possible to satisfy equation (6) over the entire frequency range. Admittedly, a variation of only 30% in ω is sufficient to encompass

half the power in the spectrum of the 5-min modes, and such a variation in the splitting cannot convincingly be ruled out by the published data¹. Nevertheless, in anticipation that cleaner results will be obtained in the future, I shall simplify the analysis by ignoring the second integral. Moreover, I shall choose a field amplitude $B(r)$ to render the first integral only weakly dependent on ω .

In the core, the major variation of ξ is produced by the Bessel function. Its argument is proportional to ω , and, in particular, the radius of the first node is roughly inversely proportional to ω . Since B is independent of ω , and ξ and B occur in the expression for $\delta_m\omega$ only in the combination ξB , it follows that $\delta_m\omega$ is stationary with respect to ω when ξ and B overlap maximally, in an appropriate sense. Therefore the variation of $\delta_m\omega$ with ω can be made weak if B and ξ are similar functions at the median frequency $\bar{\omega}$. As an example, I choose

$$B(r) = \begin{cases} B_0 \rho c u^{-1/2} J_{5/2}(\bar{\omega} u), & \bar{\omega} u \leq j_{5/2} \\ 0, & \bar{\omega} u \geq j_{5/2} \end{cases} \quad (13)$$

with $\bar{\omega}/2\pi = 3$ mHz and B_0 constant. Here j_ν is the first zero of J_ν . The factors in front of the Bessel function were chosen just for analytical convenience. The function $B(r)$ has a single maximum, near $r = 0.14R$, and falls to zero at $r \approx 0.22R$.

It is a straightforward matter to evaluate the frequency shifts $\delta_m\omega$, and hence compute B_0 by insisting that the constraint (6) be satisfied at $\omega = \bar{\omega}$. The result is such that the maximum value of $B(r)$ is 1.1×10^8 G, which implies a total field energy of 3.5×10^{45} erg. This is to be compared with a similar field suggested by Dicke³, which had been chosen to explain his oblateness data. That field extends to $r \approx 0.33R$ and has a maximum value of 6.1×10^7 G at $r \approx 0.16R$. Dicke does not exhibit the radial dependence of his field precisely, but if it were taken to be homologous to $B(r)$, the energy would be 3.6×10^{45} erg. Such a field is apparently incompatible with the oblateness data of Hill *et al.*²⁸.

The stationary value of the first integral in equation (9), with respect to variations in ω , is a minimum. Consequently, amongst function B generated from equation (13) by varying $\bar{\omega}$ and B_0 such as to preserve equation (6), 1.1×10^8 G is a maximum of the field strength.

Consider now the quadrupole modes. With respect to the coordinate system in S , only modes with $m = 0$ and $m = \pm 2$ should be detectable in the whole-disk data. As with the dipole modes, the frequencies are split by rotation. Furthermore, the magnetic field splits the frequencies for each value of m in to three, corresponding to the three different values of $|m'|$. Consequently, there are nine, not five potentially observable components in the power spectrum. Table 2 lists the synodic deviations of those nine frequencies from ω_0 , computed with that value of B_0 that reduces the dipole frequencies to a triplet. Note that there are three pairs of frequencies that are too close to have been resolved by Claverie *et al.* Thus there should be no more than six observable components, with median frequencies 0.91, 1.62, 2.39, 3.09, 3.85, 4.62 μ Hz above ω_0 . These frequencies are nearly uniformly distributed, with a mean spacing of 0.74 μ Hz.

The coincidences of some of the quadrupole frequencies arise because the magnetic splittings $\delta_m\omega - \delta_{m'-1}\omega$ of the modes are roughly integral multiples of the dipole splitting. The difference between the dipole and quadrupole eigenfunctions ξ is not very great in the core, so the integral of $[d(r\xi B)/dr]^2$ in particular is similar for the two values of l . That integral exceeds the integral of $\xi^2 B^2$ by a factor of ~ 10 . Therefore the near commensurability of the splittings is largely a product of the properties of the spherical harmonics, and is not dependent on the detailed structure of the function $B(r)$. Thus if the magnetic field in the solar core is intense enough to produce observable splitting, and is such as to reduce the expected dipole quadruplet to a triplet, then it is likely that the quadrupole frequencies constitute only a sextuplet, with a mean separation that is roughly the same as for the dipole triplet.

I have shown that some aspects of the Birmingham data might be rationalized if the Sun were to have an intense magnetic core; so might the Princeton oblateness data. However, that does not imply that the magnetic core exists. Aside from the contradiction by the oblateness measures of Hill *et al.*³⁰, there would be disturbing conflicts with certain properties of the oscillations. These may eventually be decisive. The incorrect ratios of the amplitudes of the dipole triplet, for example, would demand explanation, as would the absence of a component of the predicted quadrupole sextuplet. In superposed spectra such as those presented by Claverie *et al.*¹, it is hardly likely that some peaks are removed by interference. Nevertheless, there is the possibility that a magnetic field configuration rather different from that considered here could produce a quintuplet. Notice that in that case the splitting would still be $2\beta\Omega$, implying $\tilde{\Omega} \approx \Omega(R)$.

The high energy in the magnetic field would also need explaining. That energy is probably 50 or more times greater than the kinetic energy in the solar rotation today, though when the intense field was presumed to have been created the rotational energy may have been more than 1,000 times greater than its present value. Nevertheless, it would still be necessary to demonstrate that the regeneration of magnetic field by the differential rotation of the core, if it occurs, is sufficient to balance dissipation by instabilities.

And finally, what of the rotation itself? The potential explanation of the dipole triplet discussed here requires that the mean solar angular velocity, weighted by the reciprocal of the sound speed, is hardly different from the surface value. That is not easy to reconcile with the 12.2-day period of the core required by Dicke to explain the periodic component in his oblateness data. Moreover, it appears to be contradicted by the analyses of apparent rotational splitting inferred from oscillations in the limb-darkening function^{31,32}, which suggests that Ω is some six times greater at the centre of the Sun than it is at the surface. The condition $\tilde{\Omega} \approx 2\Omega(R)$ claimed by Claverie *et al.*¹ and Isaak² is not inconsistent with that result. Thus it seems likely that the solar magnetic field is not strong enough to have produced the splitting required to satisfy equation (6). Therefore the explanation for the fine structure in the whole-disk power spectrum needs still to be found and demonstrated, and until that is accomplished one cannot be confident that the separation of the multiplets results solely from rotational splitting.

Received 26 March; accepted 4 June 1982.

- Claverie, A., Isaak, G. R., McLeod, C. P. & van der Raay, H. B. *Nature* **293**, 443–445 (1981).
- Isaak, G. R. *Nature* **296**, 130–131 (1982).
- Dicke, R. H. *Proc. Informal Conf. on Status and Future of Solar Neutrino Research* Vol. 2 (ed. Friedlander, G.) 109–143 (Brookhaven National Laboratory, Upton, 1978).
- Dicke, R. H. *New Scientist* **83**, 12–14 (1979).
- Dicke, R. H. & Goldenberg, H. M. *Astrophys. J. Suppl.* **27**, 131–182 (1974).
- Dicke, R. H. *Phys. Rev. Lett.* **37**, 1240–1242 (1976).
- Dicke, R. H. *Solar Phys.* **47**, 475–515 (1976).
- Dicke, R. H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1309–1312 (1981).
- Dicke, R. H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1989–1993 (1981).
- Grec, G., Fossat, E. & Pomerantz, M. *Nature* **288**, 541–544 (1980).
- Fossat, E., Grec, G. & Pomerantz, M. *Solar Phys.* **74**, 59–63 (1981).
- Deubner, F.-L. *Nature* **290**, 682–683 (1981).
- Christensen-Dalsgaard, J. & Gough, D. O. *Nature* **288**, 544–547 (1980).
- Christensen-Dalsgaard, J. *Proc. 5th Eur. Meet. in Astronomy* (ed. Ledoux, P.) A.1.1–A.1.28 (Institut d'Astronomie, Liège, 1980).
- Scuflaire, R., Gabriel, M. & Noels, A. *Astr. Astrophys.* **99**, 39–42 (1981).
- Christensen-Dalsgaard, J. & Gough, D. O. *Astr. Astrophys.* **104**, 173–176 (1981).
- Christensen-Dalsgaard, J. & Gough, D. O. *Mon. Not. R. astr. Soc.* **198**, 141–171 (1982).
- Vandakurov, Yu. V. *Astr. Zh.* **44**, 786–797 (1967).
- Gough, D. O. *Mon. Not. R. astr. Soc.* **196**, 731–745 (1981).
- Brookes, J. R., Isaak, G. R. & van der Raay, H. B. *Mon. Not. R. astr. Soc.* **185**, 1–17 (1978).
- Scherrer, P. H., Wilcox, J. M., Christensen-Dalsgaard, J. & Gough, D. O. *Solar Phys.* (in the press).
- Ledoux, P. & Simon, R. *Ann. Astr.* **20**, 185–195 (1957).
- Goossens, M. *Astrophys. Space Sci.* **16**, 386–404 (1972).
- Goossens, M., Smeyers, P. & Denis, J. *Astrophys. Space Sci.* **39**, 257–272 (1976).
- Biront, D., Goossens, M., Cousens, A. & Mestel, M. *Mon. Not. R. astr. Soc.* (in the press).
- Mestel, L. *Q. Jl R. astr. Soc.* **12**, 402–416 (1971).
- Mestel, L. & Takhar, H. S. *Mon. Not. R. astr. Soc.* **156**, 419–436 (1972).
- Mestel, L., Nittman, J., Wood, W. P. & Wright, G. A. E. *Mon. Not. R. astr. Soc.* **195**, 979–1000 (1981).
- Dicke, R. H. *Solar Phys.* **78**, 3–16 (1982).
- Hill, H. A. *et al. Phys. Rev. Lett.* **33**, 1497–1500 (1974).
- Bos, R., Goode, P. R. & Hill, H. A. *Phys. Rev. Lett.* (submitted).
- Gough, D. O. *Nature* **298**, 334–339 (1982).

VHF Faraday polarization fluctuations and strong L-band amplitude scintillations near Appleton anomaly crests

A. DasGupta*, M. C. Lee† & J. A. Klobuchar*

* Air Force Geophysics Laboratory, Hanscom AFB, Massachusetts 01731, USA

† Regis College Research Center, Weston, Massachusetts 02139, USA

The Faraday rotation angle of a linearly polarized transionospheric very high frequency (VHF) signal near the crests of the Appleton anomaly, around the recent solar maximum period, often exhibits fast and intense fluctuations, when the amplitude of the signal shows saturated and very fast scintillations^{1,2}. These fluctuations usually occur during the post-sunset hours and decay around midnight. The occurrence pattern has been found to be similar to that of microwave amplitude scintillations at L-band. We present here certain observed characteristics of the fast polarization fluctuations and report the results of an experiment, which shows that there is a loss of correlation between fadings on the two circularly polarized component modes of the signal during the periods of VHF polarization fluctuations. Hence, the usual concept of Faraday rotation is no longer valid during these times. The above features are discussed in terms of our present knowledge of small scale irregularity structures in the equatorial ionosphere.

Equatorial ionospheric irregularities have been extensively investigated during the past decade by techniques such as rocket and satellite *in situ* measurements, radar, airglow and propagation effects on transionospheric signals³. It has been shown that the plumes of equatorial irregularities associated with depletions in background electron content can be studied in a very simple and inexpensive manner from simultaneous measurements of amplitude scintillations and Faraday rotation of a VHF signal propagating through the ionosphere^{4,5}. While the occurrence of depletions of the integrated electron content is commonly observed with Faraday rotation records obtained near the magnetic equator^{6,7}, a new feature, fast and intense fluctuations of the Faraday rotation angle, has recently been reported from stations situated near the crests of the Appleton anomaly^{1,2}, a region characterized by severe amplitude scintillations during solar maximum years^{8,9}. Polarization fluctuations of 136-MHz signals have been observed at Ascension Island (31°S dip), which is situated close to the southern crest of the equatorial anomaly in the African zone, and at Calcutta (32°N dip) and Ahmedabad¹⁰ (34°N dip) near the northern crest of the anomaly in the Indian sector. These have sometimes been observed even at Delhi (40°N dip) (Y. V. Somayajulu, personal communication) near the northern edge of the anomaly, but no such phenomenon has been observed at Arequipa (9°S dip), Peru or at Natal (10°S dip), Brazil which are close to the magnetic equator. Kaushika and deMendonca¹¹ possibly observed the same type of fluctuations in the polarization records of 137-MHz signals recorded at Sao Paulo (26°S dip) which is also situated near the crest of the anomaly, but they had no amplitude scintillation measurements. In their attempt to associate the fluctuations with atmospheric gravity waves, they were concerned with slower fluctuations of period 2 min and longer.

Figure 1 shows a typical polarimeter record obtained during post-sunset hours at Ascension Island (7.9°S, 14.4°W, 31°S dip). The channel at the top is the amplitude of the 136-MHz transmission from the geostationary satellite SIRIO (15°W) and the two bottom channels are the Faraday polarization angle of the signal with one shifted from the other by 90°. The polarization fluctuations are often too intense to render the Faraday data useful. The station also recorded signals from the

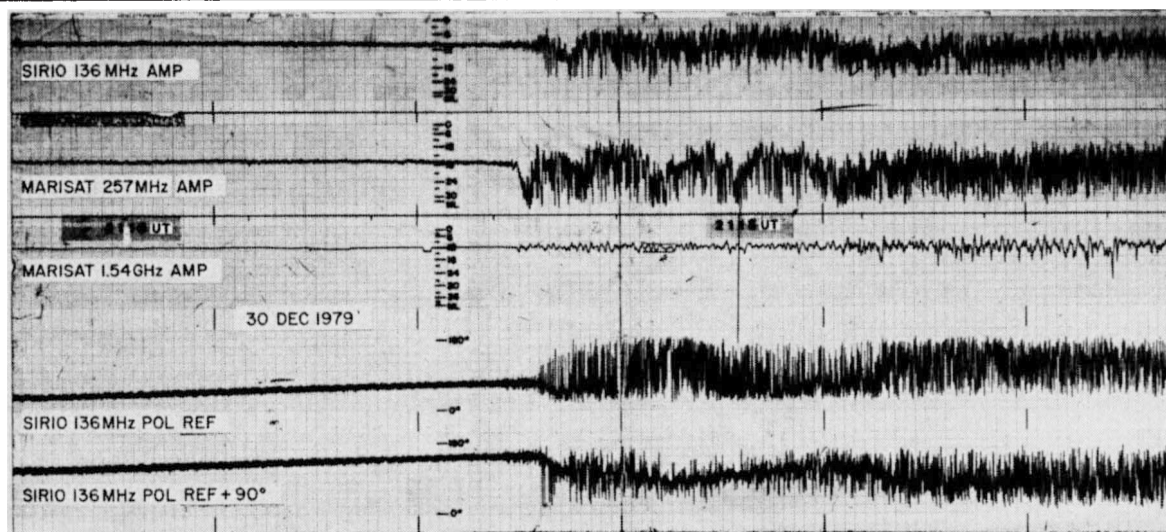


Fig. 1 Sample records of satellite signals obtained at Ascension Island. Time marks shown are in UT. Local time corresponds to UT - 1 h.

satellite Marisat (also at 15° W) at 257 MHz and 1.54 GHz for amplitude scintillation measurements. The sub-ionospheric points of the two satellites are essentially identical. The third channel from the top shows amplitude scintillations of the 1.54-GHz signal. Polarization fluctuations are found when the amplitude of the VHF signal shows saturated and very fast scintillations, which generally occur during the pre-midnight hours. We have found during more than 2 yr of observations around the recent solar maximum period that the polarization fluctuations coexist with intense *L*-band scintillations at 1.54 GHz. During the post-midnight period polarization fluctuations decay, even though VHF amplitude scintillation remains saturated, but with slower fading rates.

The change in fading rates of VHF amplitude scintillations, with and without polarization fluctuations, is clearly evident in the power spectra of 136-MHz scintillation measurements as presented in Fig. 2a and b respectively. The corresponding power spectra obtained at 257 MHz and 1.54 GHz are also shown. The approximate corner frequencies of the power spectra are indicated by arrows. The corner frequency of a power spectrum gives a measure of the fading rate, a higher corner frequency corresponding to a faster fading rate. In the present case, a lowering of the corner frequency is obvious in the late phase, as shown in Fig. 2b, when no polarization fluctuations are observed.

Power spectrum analysis of scintillation measurements is generally used to obtain information on the characteristics of the ionospheric irregularities responsible for scintillations. However, the interpretations of the power spectrum in conditions of weak and strong scintillations involve different scattering mechanisms by the ionospheric irregularities. According to the theory of radiowave propagation, for weak scintillation conditions, the amplitude fading rate of a transionospheric signal from a geostationary satellite is proportional to ν/l_F , where ν is the drift velocity of the irregularities and l_F the Fresnel dimension⁶. Because $l_F \propto \sqrt{\lambda}$ (where λ is the radiowave length), the fading rate should increase with observing frequency in weak scatter conditions. In contrast, in the case of severe scintillations, the wave propagation has to be considered in the framework of multiple scatter theory and the recorded scintillation data pertain to the ground diffraction pattern, which cannot be related to the irregularity pattern in the ionosphere in a simple manner¹². The fading rate, in severe scintillation conditions, is inversely related to the autocorrelation distance of the ground pattern. The autocorrelation interval is determined by the drift velocity as well as by the strength of scattering. As the strength of scattering (in other words, scintillation index) increases, the autocorrelation distance decreases¹³, and an increase in the fading rate is observed¹⁴. The scintillation index is usually inversely related to the observ-

ing frequency. Thus, in the case of severe scintillations, with an increase of the observing frequency the autocorrelation distance should increase and hence, a slowing down of the amplitude fading rate occurs.

In the equatorial region amplitude scintillations observed at 136 MHz and 250 MHz are usually severe (saturated or near saturated) for most of the time the phenomena are observed. Near the crests of the Appleton anomaly, severe scintillations in excess of 20 dB are quite frequent even at *L*-band (such as 1.54-GHz observations at Ascension Island) during the pre-midnight hours, particularly around the recent solar maximum⁸. The power spectra depicted in Fig. 2 are representative of scintillations observed at the corresponding frequencies in the developed and late phases of equatorial night-time ionospheric irregularities respectively.

During an intensive campaign of measurements in December 1979, in order to investigate the nature of fast polarization fluctuations, the left and right circular components of the 136-MHz signal were recorded separately in addition to the routine observations. The data were recorded on a magnetic tape in analog form, which was digitized at an 18 samples s^{-1} rate for later analysis. The cross-correlation between the amplitude fadings of the two component modes was computed over 3-min intervals. The scintillation index S_4 (refs 15, 16), which is a measure of the r.m.s. amplitude fluctuations about the mean level, is also computed for each 3 min interval at different observing frequencies. Figure 3 shows the variation of the cross-correlation coefficient between the amplitude fadings of the two circularly polarized component waves at 136 MHz with changes in the scintillation index S_4 observed at 1.54 GHz.

The correlation coefficient decreases as S_4 increases and attains low values (<0.2) when S_4 is 0.5 or more. The intervals with $S_4 > 0.5$ at 1.54 GHz generally exhibit fast polarization fluctuations at 136 MHz. In other words, the two circularly polarized components undergo quasi-independent fluctuations when strong *L*-band and polarization fluctuations are obtained. If the two modes become quasi-independent, the usual concept of Faraday rotation breaks down. The close correlation of occurrence of 136-MHz polarization fluctuations and the 1.54-GHz scintillation, clearly seen in Fig. 3, implies that either irregularities of nearly the same scale size are responsible for the two phenomena or that the irregularities causing the two phenomena coexist in the ionosphere.

Lee *et al.*¹⁷ have shown that ionospheric power law density irregularities with an outer scale size in the range 50–200 m are responsible for the 136-MHz polarization fluctuations. Faraday polarization fluctuations result from differential phase shift and the differential change in the logarithmic amplitude of the two circularly polarized modes introduced by wave scattering from these short-scale density irregularities. The theory

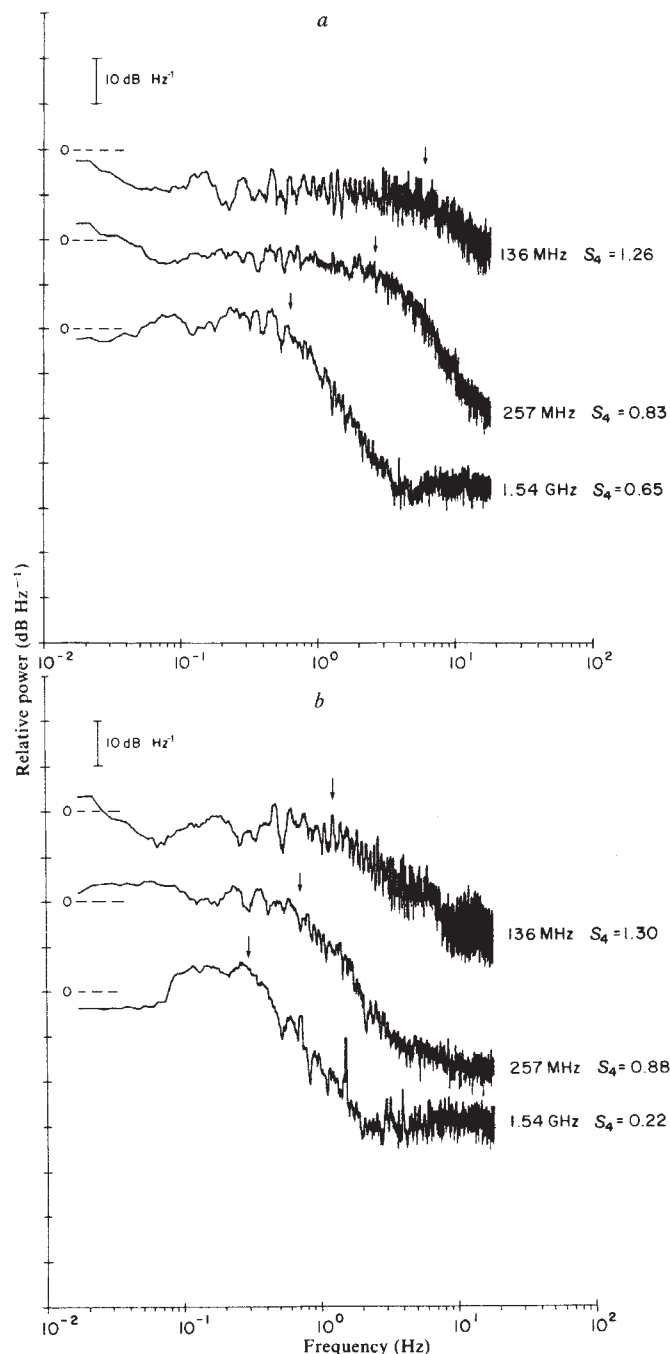


Fig. 2 Power spectra of amplitude scintillations at different frequencies for a 3-min sample centred on the time indicated. *a*, 21 h 46 min 30 s UT; Faraday polarization channels show intense and rapid fluctuations during this period. *b*, 22 h 34 min 30 s UT; no fast polarization fluctuation was observed during this period. The approximate corner frequencies are given by arrows.

shows that a deviation of $\sim 1\text{--}2\%$ in a high ambient ionization density, say, not less than $3.0 \times 10^{12} \text{ m}^{-3}$, corresponding to a plasma frequency of over 15 MHz, is required to provide a favourable environment for the 136-MHz fluctuations to occur. When the r.m.s. differential phase departure exceeds ~ 0.5 rad, the two circularly polarized modes are expected to become practically uncorrelated. Scintillations at 1.54 GHz are most effectively produced by irregularities with scale sizes of a few hundred metres (~ 300 m in the present case), which are of the same order as the Fresnel zone dimension. Both the 136-MHz polarization fluctuations and the 1.54-GHz scintillations involve density irregularities of a few hundred metres scale size. If the irregularity power spectrum follows a power law, with

diminishing spectral density at shorter scale lengths, and if the overall strength of the irregularities is high enough to have spectral density in the 50–200 m scale size range sufficient to cause polarization fluctuations at 136 MHz, it is implied that irregularities at a ≥ 300 m scale size range will also have enough spectral power to produce appreciable amplitude scintillations at 1.54 GHz. Thus, the presence of fast polarization fluctuations at 136 MHz may be taken as an indication of the possible scintillations at *L*-band and lower frequencies. The reverse may not, however, always be true.

From multi-technique observations Basu *et al.*¹⁸ have reported that, while during the pre-midnight period the power spectra of equatorial irregularities follow a power law, during post-midnight hours the overall strength of the irregularities is eroded and there is a transition to a quasi-gaussian type of spectrum indicating the decay of shorter-scale (typically < 1 km) irregularities. As the 1.54-GHz scintillation became weak, with the decay of short-scale density irregularities, the two circularly polarized modes were seen to have high cross-correlation coefficients, reflecting no, or very small, polarization fluctuations.

Except in the HF band (for example, 20 MHz), the polarization fluctuations of satellite signals cannot be effectively caused by irregularities with quasi-gaussian spectra¹⁹. Near the Appleton anomaly crests, in solar maximum conditions, the ambient ionization density (n) is very high and the high level persists during the evening hours. As a result, the absolute value of the irregularity intensity (Δn) is also very high in the anomaly region⁸. In contrast, in regions near the magnetic equator, the relatively small n and Δn and the relatively large angles of propagation with respect to the geomagnetic field make it impossible to give rise to significant polarization fluctuations of 136-MHz signals¹⁷.

We thank S. Basu, Sunanda Basu and C. L. Rino for discussions. The work at the Regis College Research Center was supported by AFGL contract F19628-80-C-0016. A.D.G. is a NRC–NAS Senior Resident Research Associate on leave from the University of Calcutta, India.

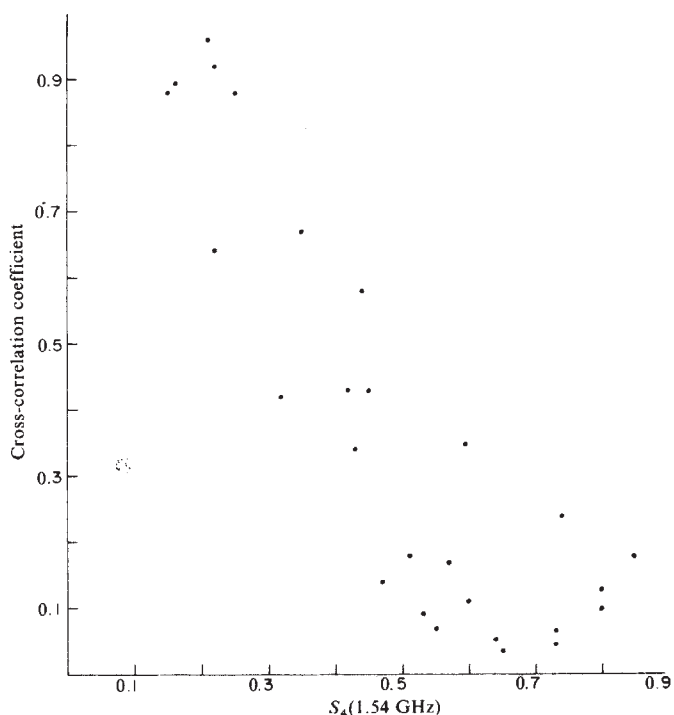


Fig. 3 Variation of cross-correlation coefficients of the fluctuations of the two opposite circularly polarized components of the 136-MHz signal with the amplitude scintillation index S_4 at 1.54 GHz. The intervals $S_4 > 0.5$ correspond to intense and fast Faraday polarization fluctuations.

Received 3 March; accepted 20 May 1982.

1. Klobuchar, J. A. & Aarons, J. *Alt. Frequenza* **49**, 345–349 (1980).
2. DasGupta, A. & Maitra, A. *Adv. Space Expl. (COSPAR Ser.)* **8**, 209–212 (1980).
3. Basu, S. & Basu, S. *J. atmos. terr. Phys.* **43**, 473–489 (1981).
4. Klobuchar, J. A., Aarons, J., Weber, E., Lucena, L. & Mendillo, M. *Paper G.4* (Radio Science Meeting, US National URSI, Boulder, Colorado, 1978).
5. Yeh, K. C., Soicher, H., Liu, C. H. & Bonelli, E. *Geophys. Res. Lett.* **6**, 473–475 (1979).
6. Yeh, K. C., Soicher, H. & Liu, C. H. *J. geophys. Res.* **84**, 6589–6594 (1979).
7. DasGupta, A., Aarons, J., Klobuchar, J. A., Basu, S. & Bushby, A. *Geophys. Res. Lett.* **9**, 147–150 (1982).
8. Aarons, A., Whitney, H. E., MacKenzie, E. & Basu, S. *Radio Sci.* **16**, 939–945 (1981).
9. DasGupta, A., Maitra, A. & Basu, S. *Radio Sci.* **16**, 1455–1458 (1981).
10. Deshpande, M. R. *et al. Curr. Sci.* **50**, 221–222 (1981).
11. Kaushika, N. D. & de Mendonca, F. *Planet. Space Sci.* **22**, 1331–1334 (1974).
12. Yeh, K. C., Liu, C. H. & Youakim, M. Y. *Radio Sci.* **10**, 97–106 (1975).
13. Rino, C. L. & Owen, J. *Radio Sci.* **16**, 31–33 (1981).
14. Umeki, R., Liu, C. H. & Yeh, K. C. *J. geophys. Res.* **82**, 2752–2760 (1977).
15. Briggs, B. H. & Parkin, I. A. *J. atmos. terr. Phys.* **25**, 339–365 (1963).
16. Whitney, H. E., Aarons, J. & Malik, C. *Planet. Space Sci.* **17**, 1069–1073 (1969).
17. Lee, M. C., DasGupta, A., Klobuchar, J. A., Basu, S. & Basu, S. *Radio Sci.* **17**, 399–409 (1982).
18. Basu, S., McClure, J. P., Basu, S., Hanson, W. B. & Aarons, J. *J. geophys. Res.* **85**, 5119–5130 (1980).
19. Lee, M. C. *J. geophys. Res.* **87**, 751–755 (1982).

Is the olivine–spinel phase transformation martensitic?

P. J. Vaughan & H. W. Green II

Department of Geology, University of California, Davis, California 95616, USA

R. S. Coe

Earth Sciences Board, University of California, Santa Cruz, California 95060, USA

Olivine, the dominant mineral in the Earth's uppermost mantle, undergoes one, or possibly a series of phase transformations to adopt a spinel structure at pressures above ~13 GPa. The mechanism by which the transformation is accomplished could have important implications for understanding mantle dynamics and deep focus earthquakes. Recently, it has been proposed that the transformation may proceed by a martensitic mechanism^{1–5}. We discuss here the characteristic properties of martensitic transformations and present evidence that the analogous transformation in Mg_2GeO_4 does not proceed by such a process.

Polymorphic transformations which occur by a martensitic mechanism involve a specific, coordinated movement of atoms. The interface which effects the transformation propagates by the movement of a series of partial dislocations resulting in the production of stacking faults. Such faults represent monolayers of the new phase⁶. This process contrasts with the nucleation and growth mechanism in which the interface propagates by a generally uncoordinated, thermally activated movement of atoms. Martensitic transformations involve a significant shear component in their transformation strain. This means that if the transformation is accomplished under nonhydrostatic stress, the process should be enhanced in favourably oriented parent crystals and retarded in unfavourably oriented parents. Higher stresses (and hence strain rates) will favour the transformation. The daughter phase generally grows as lamellae in the parent, whereas the nucleation and growth mechanism usually produces irregular interfaces due to more rapid migration in directions of greater stored strain energy in the deformed parent. Martensitic mechanisms always exhibit specific crystallographic relationships between parent and daughter crystals, whereas nucleation and growth (also proposed for olivine–spinel⁷) can, but normally does not, produce such relationships. The topotactic relation required by the martensitic mechanism proposed for the olivine–spinel transformation is $[100]_{\text{ol}}$ parallel to $\langle 111 \rangle_{\text{sp}}$ and $[001]_{\text{ol}}$ parallel to $\langle 011 \rangle_{\text{sp}}$ (refs 3,4). This crystallographic relationship has been reported in a few instances in Fe_2SiO_4 (ref. 3) and Ni_2SiO_4 (ref. 5); in the latter case only 3

out of 10 spinel crystals showed the exact relationship, and 5 showed large deviations. The occasional presence of topotaxy is compatible with any transformation mechanism, but it must be present in all instances to support the martensitic mechanism.

The olivine–spinel transformation for the chemical composition appropriate for the Earth's mantle, $(\text{Mg}_{0.9}, \text{Fe}_{0.1})_2\text{SiO}_4$, has been very difficult to study directly due to the very high transformation pressure^{8,9}. Consequently, the transformation has been studied in other systems, principally Fe_2SiO_4 (ref. 3), Ni_2SiO_4 (ref. 5) and Mg_2GeO_4 (ref. 10), which undergo the transformation at lower pressures. Although systematic studies of these and other analogues have shown very great similarities between the various systems, one complication remains. Each of the analogue systems listed above undergoes the transition directly, whereas for the Mg-rich silicate olivine system there is another (β) phase with a distorted-spinel structure which is found at intermediate pressures^{8,9}. Transmission electron microscopy (TEM) studies^{1,2,11} indicate that in many cases the β -phase may be a quench product formed at the expense of the γ -phase. Considering these uncertainties, the simplest approach regards the direct transformation (olivine–spinel) as relevant to the mantle¹². The martensitic mechanism was proposed on this basis and the topotactic relations found in Fe_2SiO_4 and Ni_2SiO_4 have been cited as support for the martensitic mechanism on the same basis. We continue this approach here while recognizing that the question of relevance to the mantle remains unresolved until the mineralogy is better known.

To test the martensitic hypothesis, we selected a deformed Mg_2GeO_4 –olivine sample which had partially transformed to spinel. The run conditions of this creep experiment (no. 92) were: confining pressure, $P = 1.3$ GPa; average differential stress, $\sigma = 280$ MPa; total axial strain 5%; duration 55 h (ref. 11). The temperature along the 18.4-mm long sample varied from 950 °C at the lower end to 1,200 °C in the centre. Thus, the lower portion of the sample was well within the spinel field but the central portion was just inside the olivine field. Specimens were selected from a small region of 50–60% spinel near the lower end of the sample. TEM foils were prepared by

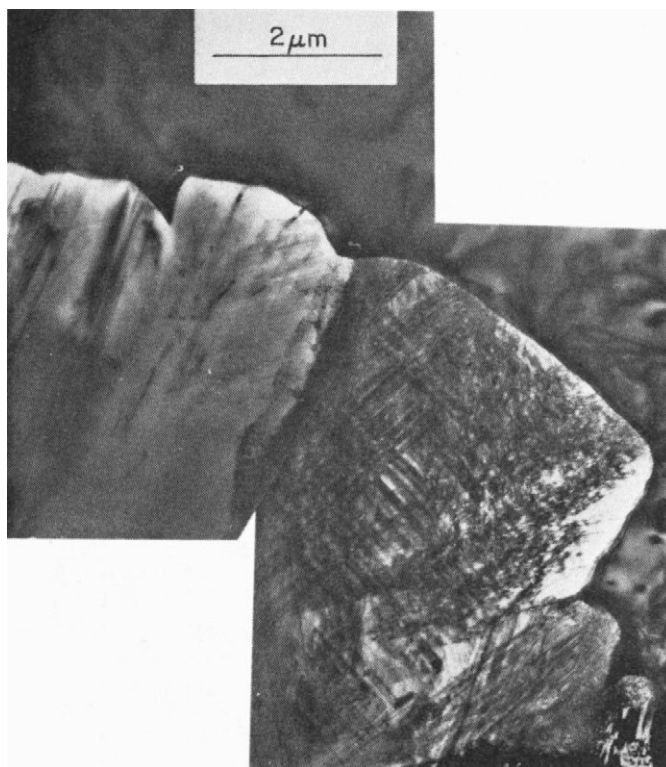


Fig. 1 High-voltage transmission electron micrograph of large spinel grains with a high density of stacking faults and parent olivine with moderate dislocation density (800 kV). The maximum compressive stress applied to the specimen lies NE–SW, parallel to the spinel–spinel grain boundaries.

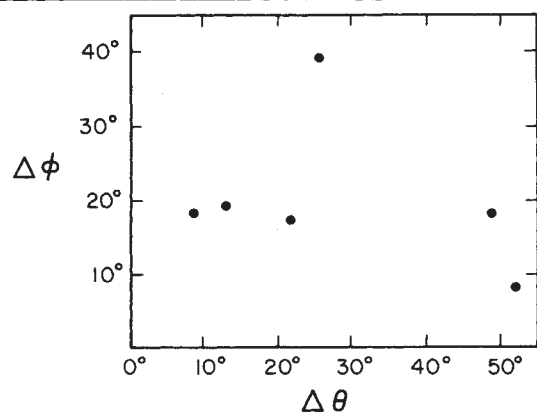


Fig. 2 Diagram showing the deviation from the orientations predicted for the martensitic transformation mechanism. $\Delta\theta$ measures the angle between $[100]_{ol}$ and the nearest $\langle 111 \rangle_{sp}$; $\Delta\phi$ is the angle between $[001]_{ol}$ and the nearest $\langle 011 \rangle_{sp}$. Data were obtained by selected area electron diffraction.

ion-thinning and examined in both optical and electron microscopes.

The following observations on this specimen suggest that the mechanism of the transformation was not martensitic:

(1) TEM images display no stacking faults in the olivine phase (Fig. 1); the substructure, both near to and far from olivine-spinel interfaces, is that normally produced during plastic deformation of germanate¹³ and silicate¹⁴ olivine. Thus, there is no direct evidence for operation of the martensitic mechanism. Stacking faults appear in the spinel phase (Fig. 1), but there is no indication that they are involved in the transformation process¹⁵.

(2) Variations in the diffraction contrast images of the daughter spinel crystals growing into the same olivine parent grain indicate that the spinels have different orientations (Fig. 1). Moreover, selected-area electron diffraction of several parent-daughter clusters demonstrates that the topotactic relations required by the martensitic mechanism are not present (Fig. 2). Small deviations from topotaxy might be explained away as loss of coherency by accumulation of dislocations in the interface. However, the lack of evidence for the required dislocations and the extremely large deviations from coherency (Fig. 2) render such an argument untenable.

(3) The spinel crystals have an elongated growth habit, but the inequian shapes are independent of the crystal orientation of both olivine and spinel. The martensitic mechanism requires a crystallographic relationship and preferential growth of spinel in olivine crystals oriented with high shear stress on the hypothetical dislocations which effect the transformation.

(4) The spinel crystals have grown preferentially in the direction of the maximum compressive stress, resulting in subparallel fingers of spinel separated by residual spikes of olivine (Figs 1 and 3).

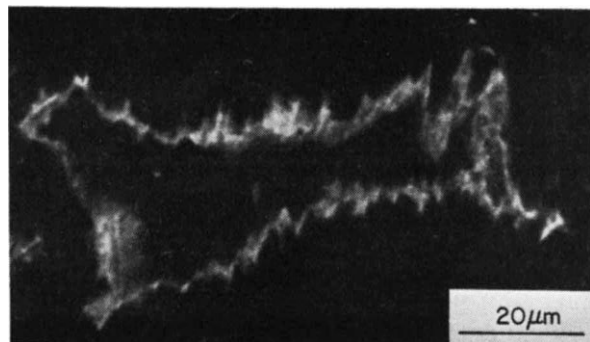


Fig. 3 Optical micrograph of a residual olivine grain surrounded by spinel. The prominent north-south spikes protruding from the olivine crystal are the remnants left between spinel grains which have grown preferentially parallel to the maximum compressive stress (crossed polarizers).

We have failed to find any evidence indicative of a martensitic transformation. On the other hand, the observation that the spinel displays a strong growth anisotropy reflecting the stress field is an interesting new discovery. This is consistent with a simple nucleation and growth mechanism in which the driving force for growth varies with the $P\Delta V$ term in the equation describing the difference in Gibbs free energy between the two phases (ref. 16, and work in preparation).

We thank J. P. Poirier and N. Hamaya for preprints and H.-R. Wenk for assistance with the electron microscopy. This work was supported by NSF EAR 74-21885A01.

Received 4 January; accepted 24 May 1982.

- Putnis, A. & Price, G. D. *Nature* **280**, 217-218 (1979).
- Price, G. D., Putnis, A. & Agrell, S. O. *Contr. Miner. Petrol.* **71**, 211-218 (1979).
- Lacam, A., Madon, M. & Poirier, J. P. *Nature* **288**, 155-157 (1980).
- Poirier, J. P. *Anelastic Properties and Related Processes in the Earth's Mantle* (AGU Monograph, in the press).
- Hamaya, N. & Akimoto, S. in *High Pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.) (Center Academic, Tokyo, in the press).
- Brooks, J. W., Loretto, M. H. & Smallman, R. E. *Acta met.* **27**, 1839-1847 (1979).
- Sung, C.-M. & Burns, R. G. *Tectonophysics* **31**, 1-32 (1976).
- Ringwood, A. E. & Major, A. *Phys. Earth planet Inter.* **3**, 89-108 (1970).
- Suito, K. *J. Phys. Earth* **20**, 225-243 (1972).
- Dachille, F. & Roy, R. *Am. J. Sci.* **258**, 225-246 (1960).
- Price, G. D., Putnis, A. & Smith, D. G. W. *Nature* **296**, 729-730 (1982).
- Poirier, J. P. *Phys. Earth planet. Inter.* **26**, 179-186 (1981).
- Vaughan, P. J. & Coe, R. S. *J. geophys. Res.* **86**, 389-404 (1981).
- Green, H. W. in *Electron Microscopy in Mineralogy* (eds Wenk, H. R. et al.) 443-464 (Springer, Heidelberg, 1976).
- Vaughan, P. J. & Kohlstedt, D. L. *Phys. Chem. Miner.* **7**, 241-245 (1981).
- Green, H. W. & Vaughan, P. J. *EOS* **62**, 1030 (1981).

Noise in chaotic systems

John Guckenheimer

Division of Natural Sciences, University of California, Santa Cruz, California 95064, USA

Procedures which estimate the noisiness of complicated, aperiodic data are outlined here and considered in the context of transitional fluid flow, although similar questions arise in many other fields^{1,2}. Because this is the initial formulation of these ideas only qualitative results are presented; any attempt to develop quantitative measures of randomness will be postponed until more testing and refinement of the procedures have been completed.

Consider a fluid experiment such as Rayleigh-Bénard convection or the flow between concentric cylinders in a regime which produces aperiodic flow³. Substantial theoretical and experimental attention has been given to the temporal characteristics of these flows as dynamical systems. The Landau-Hopf model of turbulence⁴ portrays turbulent flow as quasiperiodic with many independent frequencies. This view was successfully challenged by Ruelle and Takens⁵ with an argument based on the structural instability of quasiperiodic attractors for dynamical systems. Extensive high-precision experiments have demonstrated that the transition to aperiodic flow occurs directly from flows whose power spectra have few independent frequencies^{6,7}. These measurements for small aspect ratio fluid experiments are evidence that fluid transitions in this regime are described reasonably well by the bifurcations of systems of differential equations with few degrees of freedom⁸. There is still disagreement, however, as to whether the aperiodicity of the post-transition flows has the characteristics which one finds in a trajectory of a system of ordinary differential equations with a state space of low dimension. In particular, it has been suggested that some aspects of the data are described much better by stochastic models. The methods proposed here may help resolve this matter.

Dynamical systems with few degrees of freedom can have 'strange attractors': bounded sets of aperiodic trajectories which are asymptotically stable^{10,11}. Some features of a trajectory in

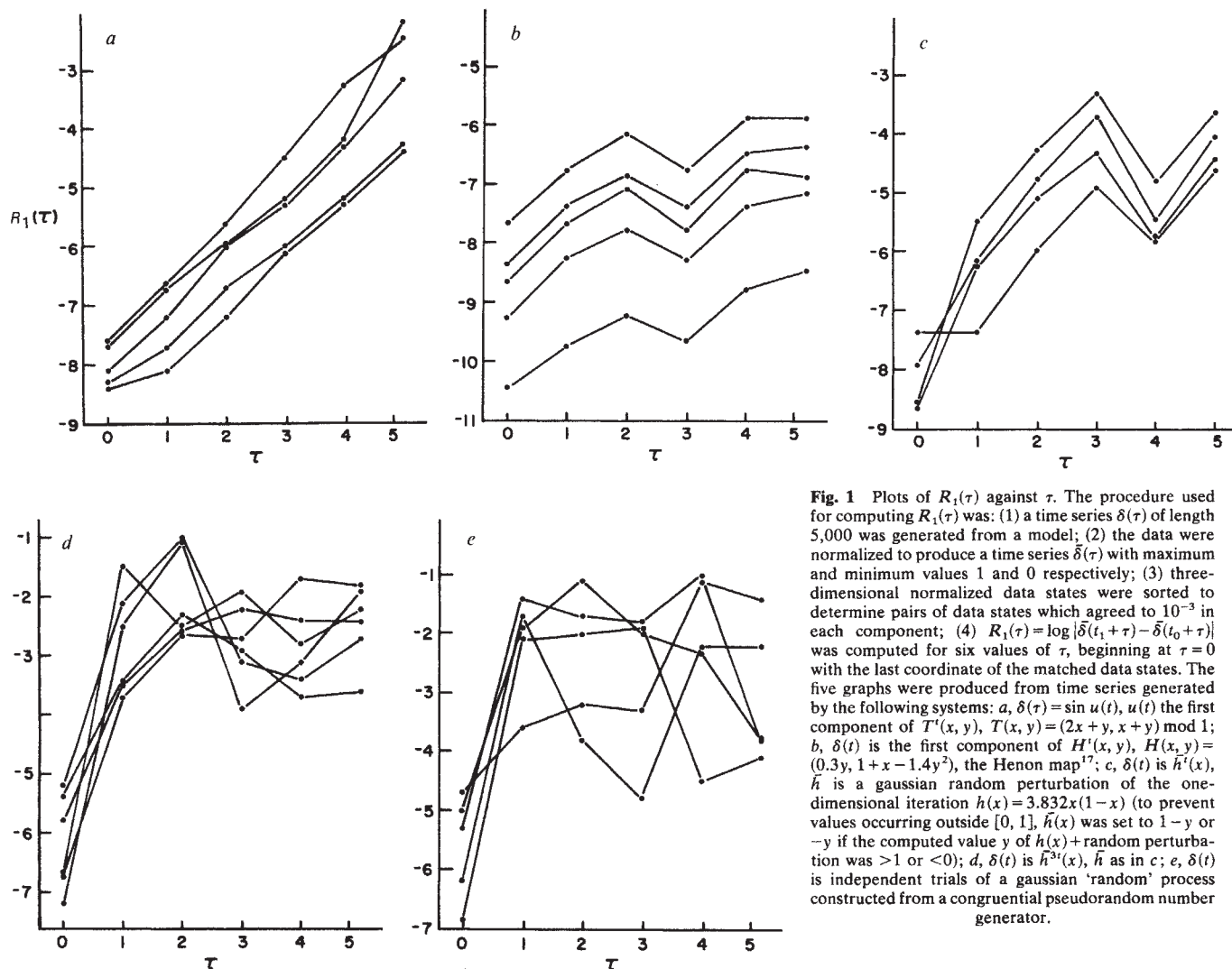


Fig. 1 Plots of $R_1(\tau)$ against τ . The procedure used for computing $R_1(\tau)$ was: (1) a time series $\delta(\tau)$ of length 5,000 was generated from a model; (2) the data were normalized to produce a time series $\tilde{\delta}(\tau)$ with maximum and minimum values 1 and 0 respectively; (3) three-dimensional normalized data states were sorted to determine pairs of data states which agreed to 10^{-3} in each component; (4) $R_1(\tau) = \log |\tilde{\delta}(t_1 + \tau) - \tilde{\delta}(t_0 + \tau)|$ was computed for six values of τ , beginning at $\tau = 0$ with the last coordinate of the matched data states. The five graphs were produced from time series generated by the following systems: *a*, $\delta(\tau) = \sin u(t)$, $u(t)$ the first component of $T'(x, y)$, $T(x, y) = (2x + y, x + y) \bmod 1$; *b*, $\delta(t)$ is the first component of $H'(x, y)$, $H(x, y) = (0.3y, 1 + x - 1.4y^2)$, the Henon map¹⁷; *c*, $\delta(t)$ is $\tilde{h}(x)$, $\tilde{h}$ is a gaussian random perturbation of the one-dimensional iteration $h(x) = 3.832x(1 - x)$ (to prevent values occurring outside $[0, 1]$, $\tilde{h}(x)$ was set to $1 - y$ or $-y$ if the computed value y of $h(x) + \text{random perturbation}$ was > 1 or < 0); *d*, $\delta(t)$ is $\tilde{h}^{3t}(x)$, $\tilde{h}$ as in *c*; *e*, $\delta(t)$ is independent trials of a gaussian 'random' process constructed from a congruential pseudorandom number generator.

a strange attractor appear chaotic even though its evolution is determined completely by initial conditions. There is a sensitive dependence to initial conditions which prevents long-term prediction of the location of a trajectory within the attractor due to the uncertainty in our measurements of initial conditions. I interpret the Ruelle-Takens theory as asserting that aperiodic fluid flow is described by a strange attractor. This behaviour contrasts with randomness in the flow. Here randomness is defined as unpredictability in the short term evolution of a system. The present aim is to estimate the extent of randomness or noise in a system from the data which it produces.

In analysing a time series, I begin with the hypothesis that it can be described by evaluating a function along a trajectory in a strange attractor of a system with few degrees of freedom. This hypothesis implies that the trajectory will visit some regions of the state space repeatedly, returning many times to positions that are close to ones that it occupied previously. When this happens, the corresponding observations in the time series approximate each other. Indeed, entire sections of the time series approximate each other over time intervals for which the trajectories remain close. A second assumption about the function defining the time series allows this observation to be reversed.

Packard *et al.*^{12,13} and Takens¹⁴ have developed a procedure for recovering the location of a point in the state space from a section of the time series based on the Whitney embedding theorem¹⁵. Specifically, the trajectory with initial condition x is denoted by $\phi(x, t)$ and a time interval Δt selected. We then

assume that the functions f_i defined by $f_i(x) = f(\phi(x, i\Delta t))$ are in a general position. The embedding theorem then implies that the mapping F defined by $F(x) = (f_0, f_1, \dots, f_k)$ is an embedding of the state space into R^{k+1} if k is at least twice the dimension of the state space. We need only observe here that if $\phi(x, t)$ is a trajectory, then the observations $f(\phi(x, t + i\Delta t))$, $0 \leq i \leq n$, determine uniquely the location of $\phi(x, t)$ when n is large enough.

Sections of a discrete time series $\delta(t)$ are now used which approximate each other to assess the extent to which the data $\delta(t)$ could have been generated by a dynamical system with a low dimensional state space. An integer k is selected and a $(k + 1)$ -dimensional data state $\Delta(t)$ defined to be a $(k + 1)$ vector $(\delta(t), \delta(t + 1), \dots, \delta(t + k))$. If k is sufficiently large and the data can be generated by a strange attractor, then $\Delta(t + 1)$ is determined by $\Delta(t)$ by means of an unknown $(k + 1)$ order (nonlinear) difference equation.

The practical problems here are that (1) no physical process is entirely free of randomness, and (2) any experiment gives us limited quantities of data. Nonetheless, the data states defined above can be used to estimate the randomness in a system. In two procedures used as an initial test for these methods, the data states of a time series are sorted to locate efficiently the pairs and triples of data states which are close to each other. Longer sections of the data which include these data states can then be examined.

The first procedure is based on the observation that the divergence of trajectories from one another in a strange attractor is approximately exponential, at a rate determined by the

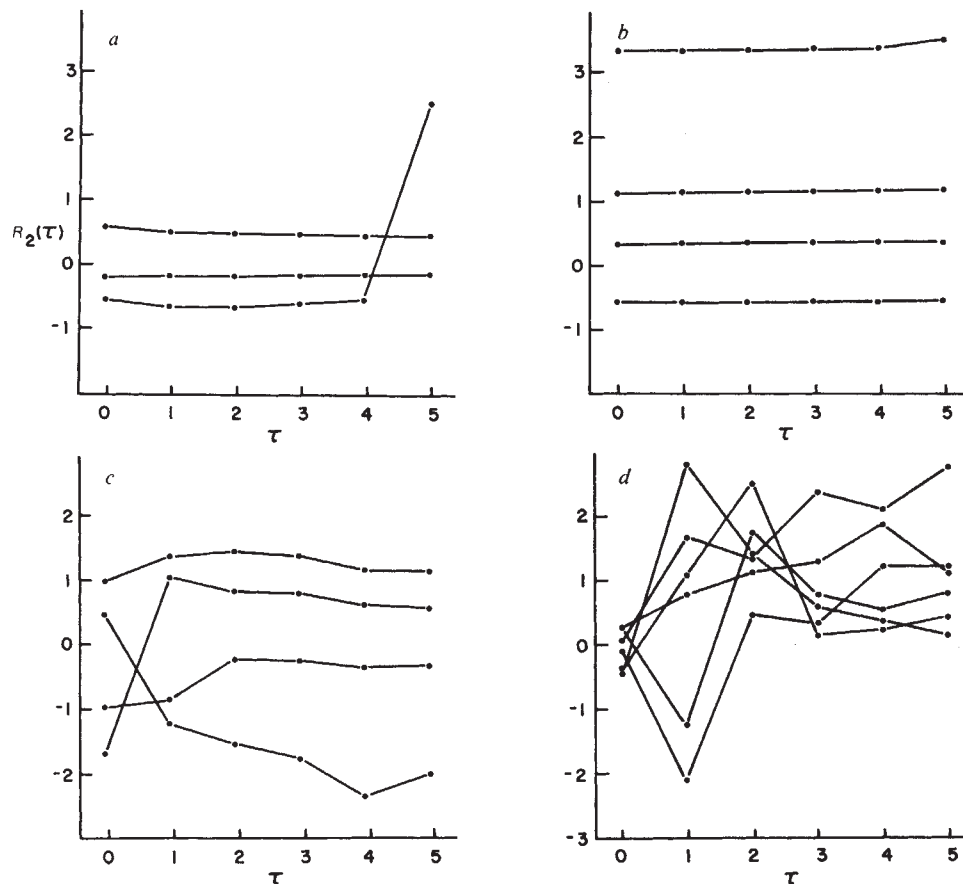


Fig. 2 Plots of $R_2(\tau)$ against τ for the same time series $\delta(t)$ used in Fig. 1a–d. To compute R_2 from $\delta(t)$, nearby triples of data states were located. With $\tau=0$ corresponding to the last component of the matched data states,

$$R_2(\tau) = \frac{\delta(t_2 + \tau) - \delta(t_1 + \tau)}{\delta(t_0 + \tau) - \delta(t_1 + \tau)}$$

was computed for $\tau = 0, \dots, 5$.

largest Liapunov exponent of the trajectory. This suggests that the difference $R_1(\tau) = \log |\delta(t_1 + \tau) - \delta(t_0 + \tau)|$ of two sections of the time series $\delta(t)$ should be approximately linear in τ with slope which is generally independent of the sections of data chosen in a given region of the state space. Thus the first step is to measure $R_1(\tau)$ for sections of data which immediately follow pairs of nearby data states $\Delta(t_0)$ and $\Delta(t_1)$.

There is a limitation in using $R_1(\tau)$ due to the nonuniformity of the spreading of trajectories in different parts of a strange attractor and to the nonlinearity of the observing function f . The second procedure examined the statistics

$$R_2(\tau) = \frac{\delta(t_2 + \tau) - \delta(t_1 + \tau)}{\delta(t_0 + \tau) - \delta(t_1 + \tau)}$$

at times following the close approach of the corresponding triple of data states. In sections of a time series for which the strange attractor hypothesis indicates exponential divergence, $R_2(\tau)$ should be almost constant. In our present tests R_2 seems to be less susceptible to the nonlinearity of the observable and the nonuniformity of an attractor than R_1 .

The statistics R_1 and R_2 are designed to examine more directly the extent of the randomness in a system than the procedures described by Takens¹⁴ and Froehling *et al.*^{12,13}. Takens partitions the space of k -dimensional data states into boxes of size ϵ . He uses the number $N(k, \epsilon)$ of boxes populated for varying k and ϵ to estimate the topological entropy and capacity of a strange attractor. However, in a system with randomness, the noise can be expected to dominate the counts $N(k, \epsilon)$ when ϵ is considerably smaller than the amplitude of the noise. This places limitations on the suitability of Takens' process for data with moderate amounts of noise. One might obtain an indication of the noise amplitude in a set of data if there is a pronounced plateau in $N(k, \epsilon)$ with varying ϵ , but the issue of short term predictability is not addressed directly. Froehling *et al.*^{12,13} use a method of local linear regression to fit the set of data states in a given region of the space of data

states to a linear subspace. This procedure is complementary to the one described here and can be used to select an optimal value for the dimension of the data states which should be used when computing the statistics R_1 and R_2 . The LLR method does not appear to be effective in assessing the issue of short term predictability.

The extreme examples which have been examined with these two procedures are a linear toral Anosov diffeomorphism¹⁶ and independent trials of a gaussian random variable. The toral diffeomorphism is defined by $T(x, y) = (2x + y, x + y)$, where x and y are regarded as cyclic coordinates (mod 1). The whole two-dimensional torus is a strange attractor for T . The nonlinear function $f(x, y) = \sin x$ was used to generate time series $\delta(t)$ for this system. Examples have also been examined of the time series drawn from the x -coordinate of iterates of the Henon mapping¹⁷ defined by $H(x, y) = (0.3y, 1 + x - 1.4y^2)$, and a random perturbation of a one-dimensional quadratic iteration $h(x) = 3.832x(1 - x)$. This perturbation was formed by adding to $h(x)$ a gaussian random variable constructed from a congruential pseudorandom number generator.

Figure 1 shows the results obtained from applying the first procedure to time series of 5,000 observations drawn from each of the examples. Nearby three-dimensional data states have been located, and then $R_1(\tau)$ has been plotted for the last components of the nearby data states and for the five observations immediately following these. The results clearly separate the deterministic systems from the pseudorandom gaussian process. The nonuniformity of the expansion in the Henon mapping is reflected in the smaller values of R_1 at the third observation following the matched data states compared to the second. For the one-dimensional iteration, the deterministic features dominate the randomness at this scale. To show the randomness more clearly, Fig. 1d shows the results of selecting every third data point (there is a stable period 3 cycle in the deterministic map) and then working on a finer scale of resolution. The randomness in the data shows clearly in Fig. 1d. Note also in Fig. 1d and e the large magnitude of the uncertainty

in the knowledge of the next observation associated with our current knowledge of the data state.

Figure 2 shows R_2 statistics for the same time series. The gaussian process produced only one triple which was a close match, and this was very erratic. Only graphs with moderate values of the R_2 statistic are shown. The R_2 statistic works strikingly well for the Anosov and Henon mappings, while the one-dimensional iteration gives an indication of a small amount of randomness. Looking at the finer scale of resolution for the third iterate of this process, the randomness is once again dominant.

The results of these initial test on models of strange attractors and stochastic systems are promising. The differences between time series generated by strange attractors and those generated by systems with some randomness are apparent. There are three specific indications of determinism versus randomness in the tests we have performed with these model data:

(1) If $\tau = 0$ denotes the time of the last observations used in

Received 1 February; accepted 28 April 1982.

1. Guckenheimer, J. *Lect. Math. Life Sci.* **10**, 1–47 (1978).
2. Brilling, D. R., Guckenheimer, J., Guttorp, P. & Oster, G. *Lect. Math. Life Sci.* **13**, 65–90 (1981).
3. Swinney, H. L. & Gollub, J. (eds) *Hydrodynamic Instabilities and the Transition to Turbulence* (Springer, New York, 1981).
4. Landau, L. & Lifshitz, E. M. *Fluid Mechanics* (Pergamon, Oxford, 1959).
5. Ruelle, D. & Takens, F. *Comm. Math. Phys.* **20**, 167–192 (1971).
6. Libchaber, A. & Maurer, J. *Nonlinear Phenomena and Phase Transitions* (ed. T. Riste) (Geilo Nato School, in the press).
7. Gollub, J. P. & Benson, S. V. *J. Fluid. Mech.* **100**, 449–470 (1980).
8. Eckmann, J.-P. *Rev. mod. Phys.* **53**, 643–654 (1981).

Natural production of tritium in permeable rocks

J. N. Andrews

School of Chemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

R. L. F. Kay

Institute of Geological Sciences, Environmental Protection Unit, Building 151, Harwell Laboratory, Didcot, Oxon OX11 0RA, UK

Natural tritium production by the reaction, ${}^6\text{Li}(n, \alpha){}^3\text{H}$, may in some circumstances lead to ${}^3\text{H}$ contents in the fracture fluids of a granite, which are significantly above the present natural ${}^3\text{H}$ content (0.5 tritium units, TU) of pre-bomb precipitation. We show here that the principal factors which determine the ${}^3\text{H}$ content of the fracture fluids are the U, Th, Li and B contents of the rock and its matrix porosity. Granites with high U, Th and Li contents; low B content and a low matrix porosity may have a naturally maintained ${}^3\text{H}$ content of up to about 2.5 TU in water within the matrix. This has important implications for the interpretation of groundwater flow patterns from their ${}^3\text{H}$ contents. Whereas it has generally been assumed that groundwaters with ${}^3\text{H}$ contents greater than 0.5 TU have a proportion of post-1954 recharge present, this may not always be a valid conclusion.

Precipitation after 1954 contained ${}^3\text{H}$ which was produced by the thermonuclear weapon testing in the atmosphere. The ${}^3\text{H}$ content of rainfall in the UK has declined from a maximum of ~2,000 TU in 1963 to about 60 TU in 1975¹. The pre-1954 ${}^3\text{H}$ content (due to cosmic ray production) of precipitation on the western coast of Europe was ~2.5 TU (ref. 2) and a groundwater which was recharged in 1953 with this ${}^3\text{H}$ content would at present (1982) contain only a residual 0.5 TU because of ${}^3\text{H}$ decay (half life, 12.3 yr). A groundwater with the natural ${}^3\text{H}$ input (2.5 TU) would need to be ~60-yr old to reduce its ${}^3\text{H}$ content to 0.1 TU by decay. This is about the analytical limit³ for determining ${}^3\text{H}$ in water.

A groundwater with a ${}^3\text{H}$ content ≥ 0.1 TU is therefore generally considered to be younger than 60 yr and, especially if the ${}^3\text{H}$ content exceeds a few TU, to contain a proportion of

matching nearby data states, then a large value of $R_1(1) - R_1(0)$ shows that one state has exceeded the resolution at which one step prediction is possible.

(2) Approximately constant differences in the values of R_1 computed from different trajectories demonstrate a consistent pattern of separation of trajectories like that present in strange attractors.

(3) Approximately constant values of $R_2(\tau)$ compared with τ give an indication similar to (2) of the orderly separation of trajectories found in strange attractors.

I suggest that statistics such as R_1 and R_2 may help resolve questions about whether the irregularities of time series drawn from experimental observations are due to effects which can be described by deterministic systems with few degrees of freedom. The problems discussed above are related to questions about entropy and fractal dimension of data¹⁴, but the methods reported here address the question of randomness as the cause of irregularity more directly.

9. Greenside, H. S., Ahlers, G., Hohenberg, P. C. & Walden, R. W. *Physica D* (in the press).
10. Guckenheimer, J., Moser, J. & Newhouse, S. *Dynamical Systems* (Birkhäuser, Boston, 1980).
11. Ruelle, D. *Ann. N.Y. Acad. Sci.* **316**, 408–416 (1979).
12. Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. *Phys. Rev. Lett.* **45**, 712–716 (1980).
13. Froehling, H., Crutchfield, J. P., Farmer, J. D., Packard, N. H. & Shaw, R. S. *Physica* **3D**, 605–617 (1981).
14. Takens, F. in *Dynamical Systems and Turbulence*, Warwick 366–381 (Springer, New York, 1981).
15. Hirsch, M. W. *Differential Topology* (Springer, New York, 1976).
16. Smale, S. *Bull. Am. math. Soc.* **73**, 747–817 (1967).
17. Henon, M. *Comm. Math. Phys.* **50**, 69–78 (1976).

post-1954 precipitation. Here, we consider whether or not it is possible that groundwaters older than 60 yr could contain levels of ${}^3\text{H} > 0.1$ TU due to *in situ* production of ${}^3\text{H}$. Such production is most likely for groundwaters from granites for which lithium and radioelement contents are comparatively high.

The only nuclear reaction which can produce significant ${}^3\text{H}$ natural activity by in-ground production is the neutron-induced reaction:



The thermal neutron cross-section for this reaction⁴ is large and neutrons are produced in rocks by both spontaneous fission of uranium and by (α, n) reactions on the major matrix elements; Si, O, Al, Mg. The activity of ${}^3\text{H}$ produced at a constant rate P by neutron irradiation of lithium for time t is given by:

$$\left(\frac{dN_t}{dt}\right)_t = \lambda(N_t)_t = P(1 - e^{-\lambda t}) \quad (2)$$

where $(N_t)_t$ is the number of ${}^3\text{H}$ atoms present after time t and λ is the decay constant for ${}^3\text{H}$, from which

$$(N_t)_t = \frac{P}{\lambda}(1 - e^{-\lambda t}) \quad (3)$$

and

$$\lim_{t \rightarrow \infty} (N_t)_t = \frac{P}{\lambda} \quad (4)$$

P may be estimated in a manner similar to that used by Zito *et al.*⁵ to estimate in-ground ${}^{14}\text{C}$ production. The chance, F , of neutron absorption by Li relative to absorption by other neutron-absorbing elements is given by:

$$F = \sigma_{\text{Li}} N_{\text{Li}} / \sum_i \sigma_i N_i = \sigma_{\text{Li}} N_{\text{Li}} / C_{\text{r}} \quad (5)$$

where σ_i , N_i are the total neutron absorption cross-sections and abundances for all elements in the matrix; σ_{Li} is the cross-section for reaction (1) and N_{Li} is the abundance of Li in the matrix. The $\sigma_i N_i$ summation, C_{r} , may be regarded as the total neutron capture probability for the rock matrix and the terms

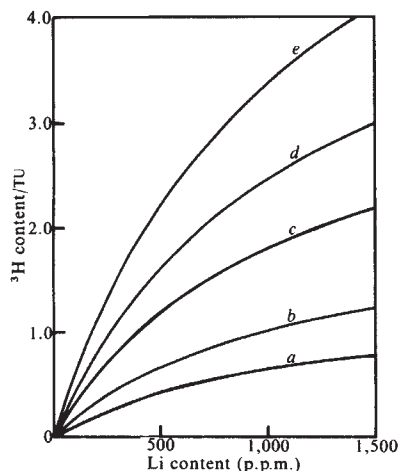


Fig. 1 Variation of the ^3H content of water in a granite with the Li, U and Th contents. A boron content of $15\ \mu\text{g per g}$ (average granite, ref. 6) and an equivalent total porosity (see text) of 1% were assumed. *a*, 3.5 p.p.m. U+18 p.p.m. Th; *b*, 10 p.p.m. U+15 p.p.m. Th; *c*, 20 p.p.m. U+20 p.p.m. Th; *d*, 30 p.p.m. U+20 p.p.m. Th; *e*, 40 p.p.m. U+30 p.p.m. Th.

of this summation for various rocks are listed in Table 1. The Li-content of the rock matrix directly influences the ^3H production rate and B, because of its strong influence on the total capture probability, needs to be further considered. This may be facilitated by substituting for C_i in equation (5) as follows:

$$C_i = C + \sigma_{\text{Li}} N_{\text{Li}} + \sigma_{\text{B}} N_{\text{B}} \quad (6)$$

where C is the total absorption probability excluding the contributions of Li and B. For a granite matrix, the fraction F , of thermal neutrons captured by Li may be shown to be:

$$F = [\text{Li}] / \{1.057 \times 10^{-7} + [\text{Li}] + 6.827[\text{B}]\} \quad (7)$$

where $[\text{Li}]$ and $[\text{B}]$ are the Li and B contents of the rock in $\mu\text{g per g}$. This equation was derived for the elemental composition of average granite⁶ using cross-sections for thermal neutron absorption. Only thermal neutron absorption has been considered as the cross-sections for fast neutron absorption reactions are generally very much smaller than those for thermal neutron reactions and the high abundance of light elements (O, Si) in the matrix ensures that neutrons are thermalized by

scattering processes. The ^3H production rate, P , is given by:

$$P = Fn \quad \text{atoms s}^{-1} \text{cm}^{-3} \quad (8)$$

where n = neutron production in the rock, neutrons $\text{s}^{-1} \text{cm}^{-3}$. The neutron production in granites has been estimated by Feige *et al.*⁷ (Table 2).

Following neutron-induced fission of ^6Li , the ^3H nuclei formed capture electrons to form ^3H atoms which diffuse through the rock matrix much more readily than molecular hydrogen. Their ultimate fate is to combine to form $^3\text{H}_2$ or to exchange with the H atoms of water molecules to form ^3HHO . The latter is much more probable than direct combination in a matrix which contains aqueous fluids as the diffusing ^3H atoms are likely to encounter water molecules before they meet other ^3H atoms. Exchange of the diffusing ^3H atoms with structurally bound water or hydroxyl groups in the rock matrix is also likely to occur. The total water content of the matrix, that is, structurally bound water and pore- or fracture-contained water, may be combined to obtain an equivalent total porosity, ϕ , for the rock. The ^3H atoms are trapped on a random basis by the total water in the matrix and its specific activity in structurally bound and pore water is therefore identical with that in the fracture fluids from which samples are generally drawn.

The ^3H content of the fluid is given by the equation:

$$^3\text{H-content(TU)} = \frac{P}{\lambda \phi \times 2 \times 6.02 \times 10^{23}} \quad (9)$$

where λ (0.0559 yr^{-1}) is the decay constant of ^3H and P is the ^3H production rate (atoms yr^{-1}); as there are $2 \times 6.02 \times 10^{23}$ H atoms in 1 mole (18g) of water and 1 TU is defined as 1 ^3H atom in 10^{18} H atoms. The ^3H content is inversely proportional to the equivalent total porosity and will become most significant for rocks with both low porosity and only small amounts of structurally bound water. Recent measurements on various granites^{8,9} suggest that the minimum porosity is in the range 0.3–0.5%. The amount of structurally bound water is also about 0.5% for unaltered granites⁹. The minimum equivalent total porosity for granite is therefore $\sim 1\%$ and *in situ* ^3H contents were calculated for this porosity. Figure 1 shows the variation of ^3H content with the Li and radioelement contents for a granite of average B content ($15\ \mu\text{g per g}$). At high Li and radioelement contents, *in situ* ^3H production could lead to ^3H contents in the fracture fluids which exceed the maximum possible residual cosmogenic ^3H from recharge just before 1954

Table 1 Neutron capture probabilities (abundance $\times$ capture cross-section) in rock matrices

Element	Atomic No.	Cross-section (barns)	Neutron capture probability (mol barns)				
			Ultramafic	Basalt	Granite	Sandstone	Limestone
Li	3	71	0.000005	0.00015	0.00040	0.00015	0.00034
B	5	755	0.00007	0.00035	0.00104	0.00244	0.00140
Na	11	0.534	0.00013	0.00045	0.00064	0.00007	0.00000
Mg	12	0.064	0.00068	0.00012	0.00000	0.00000	0.00012
Al	13	0.232	0.00004	0.00075	0.00066	0.00021	0.00004
Si	14	0.160	0.00108	0.00137	0.00184	0.00209	0.00014
Cl	17	33.6	0.00005	0.00005	0.00022	0.00014	0.00014
K	19	2.07	0.00002	0.00044	0.00176	0.00056	0.00014
Ca	20	0.44	0.00008	0.00074	0.00017	0.00042	0.00332
Ti	22	6.1	0.00004	0.00115	0.00029	0.00019	0.00005
Cr	24	3.1	0.00012	0.00001	0.00000	0.00000	0.00000
Mn	25	13.3	0.00036	0.00048	0.00014	0.00001	0.00027
Fe	26	2.56	0.00045	0.00039	0.00123	0.00044	0.00017
Co	27	37.5	0.00013	0.00003	0.00000	0.00000	0.00000
Ni	28	4.54	0.00015	0.00001	0.00000	0.00000	0.00000
Sm	62	5,820	0.00002	0.00019	0.00034	0.00038	0.00005
Gd	64	49,000	0.00016	0.00156	0.00280	0.00311	0.00040
Total capture probability			0.00358	0.00841	0.01177	0.01045	0.00668

Elements which do not contribute $>1\%$ of the total capture probability for at least one rock type, have not been tabulated. The average rock compositions of ref. 6 have been used.

Table 2 Neutron production in granites**a, Spontaneous fission (s.f.) neutrons**

	s.f. half life (yr)	Neutrons per fission	Neutrons per year per μg of natural U
^{238}U	8.04×10^{15}	2.2	0.4762
^{235}U	1.87×10^{17}	2.5	0.0002
Total	—	—	0.4764

b, Neutrons produced by α -irradiation of light nuclei⁷

SiO ₂ content (%)	Neutrons per year per g of rock for each p.p.m.	
	U	Th
60	1.70	0.76
70	1.44	0.65

(0.5 TU). The effect of increased B contents is demonstrated by Table 3. In general, a 10-fold increase in B content results in a 3- to 5-fold reduction in the amount of *in situ* produced ^3H in the fracture fluids. Variation of the Gd content of the matrix would also significantly affect ^3H production.

The Carnmenellis granite, one of the major outcrops of the granite batholith of South West England¹⁰, is known to have high radioelement and Li contents. High salinity groundwaters occur in tin mines at depths of up to 690 m, and these have been shown to result from alteration reactions of the granites with meteoric waters¹¹. These waters had apparently constant ^3H contents of 6 ± 2 TU over a range of salinities and this could be a consequence of either *in situ* ^3H production or variations in the extent to which alteration reactions have proceeded.

Lithium is generally present in the biotite¹² and an average whole-rock Li content for the Carnmenellis granite is 278 μg per g, compared with an average Li content of 340 μg per g for the similar, poorly megacrystic, Dartmoor granite¹³. Higher average Li contents ($\sim 1,100$ μg per g) are present in the St Austell granite¹².

The available data for the Carnmenellis granite suggest that its B content is of the order 250 μg per g, compared with 15 μg per g for average granite. Outcrop samples of the Bodmin granite have an average B content of 270 μg per g whilst the two available analyses (J. R. Hawkes personal communication) of the Carnmenellis granite are 85 and 230 μg per g respectively. Tourmaline is the principal B-bearing mineral and is present in both primary and secondary forms¹⁰. The variation of B content with depth in the granite is not known but it is likely that secondary tourmalinization is greater towards the granite margins. The B content at depth is therefore unlikely to be greater than that at outcrop. The Gd contents of the Cornish granites¹⁴ are similar to those for average granite.

The dependence of ^3H production in the Carnmenellis granite on Li and B contents is indicated in Table 3. It is evident that for the probable geochemical conditions in this granite (250–500 μg per g of Li; 250 μg per g of B), the *in situ* ^3H production cannot supply more than 0.5 TU to the fracture fluids. Higher Li contents occur in the nearby St Austell granite for which

the B content is lower (< 50 μg per g), so that tritium in excess of 1 TU could be present in fracture fluids as a result of *in situ* production. It is, however, concluded that the 6 TU observed in the deep groundwaters from the Carnmenellis granite must be principally due to the presence of meteoric water of post-1954 origin. In the most favourable case for ^3H production, a granite with a high Li content and an average B content, the consequent fracture fluid ^3H content is unlikely to exceed 1 TU in most cases, or up to 2.5 TU for highly uraniferous granite.

We thank A. H. Bath, W. M. Edmunds, J. R. Hawkes and M. Stone for helpful discussions, J. R. Hawkes for access to unpublished analyses of the Cornish granites by the Laboratory of the Government Chemist. The support of the Natural Environment Research Council through research grant GR3/2622 is gratefully acknowledged. R.L.F.K. contributes to this paper with the permission of the director of the Institute of Geological Sciences.

Received 8 March; accepted 17 May 1982.

1. Foster, S. S. D. & Smith-Carrington, A. *J. Hydrol.* **46**, 343–364 (1980).
2. von Buttlar, H. & Libby, W. F. *J. Inorg. nucl. Chem.* **1**, 75–91 (1955).
3. Cameron, J. F. in *Radioactive Dating and Methods of Low-level Counting* 543–573 (IAEA, Vienna, 1967).
4. Hughes, D. J. & Schwarz, R. B., *Report BNL-325* (Brookhaven National Laboratory, New York, 1958).
5. Zito, R., Donahue, D. J., Davis, S. N., Bentley, H. W. & Fritz, P. *Geophys. Res. Lett.* **1**, 235–238 (1980).
6. Parker, R. L. *U.S. geol. Surv. Profess. Pap.* 440–D (1967).
7. Feige, Y., Oltman, B. G. & Kastner, J. *J. geophys. Res.* **73**, 3135–3142 (1968).
8. Hall, D. H., Black, J. H. & Storey, B. C. *Rep. ENPU 81-10*, (Institute of Geological Sciences, London, in the press).
9. Nelson, P. *et al.*, *Rep. LBL-8280* (Lawrence Berkeley Laboratory, University of California 1979).
10. Edmonds, E. A., McKeown, M. C. & Williams, M. *British Regional Geology: South West England* (Institute of Geological Sciences, London, 1975).
11. Burgess, W. G., Edmunds, W. M., Andrews, J. N., Kay, R. L. F. & Lee, D. J. *The Origin and Circulation of Groundwater in the Carnmenellis Granite: the Hydrogeochemical Evidence* (Institute of Geological Sciences, London, 1982).
12. Beer, K. E., Edmunds, W. M. & Hawkes, J. R. *Energy* **3**, 281–292 (1978).
13. Hawkes, J. R. & Dangerfield, J., *Proc. Ussher Soc.* **4**, 158–171 (1978).
14. Alderton, D. H. M., Pearce, J. A. & Potts, J. A. *Earth planet. Sci. Lett.* **49**, 149–165 (1980).
15. Tammemagi, H. Y. & Wheildon, J. *Geophys. J. R. astr. Soc.* **38**, 83–94 (1974).

Role of Mn(IV) oxide in abiotic formation of humic substances in the environment

H. Shindo* & P. M. Huang

Saskatchewan Institute of Pedology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

The oxidative polymerization of polyphenols in soils is regarded as one of the main processes in the formation of humic substances^{1–7}. The polymerization causes the darkening of the colour of polyphenols—the ‘browning phenomenon’—which can be accelerated non-enzymatically as well as enzymatically^{7–15}. However, the relative effectiveness of various inorganic components, especially noncrystalline to poorly crystalline Fe, Al, Mn and Si oxides commonly occurring in soils^{16–19}, in affecting the browning of polyphenols is still obscure. The elucidation of the relative effectiveness of these oxides is of importance in understanding their roles in the formation of humic substances in the environment. We report here that Mn(IV) oxide is a very effective oxidant with respect to the abiotic browning of hydroquinone solution over the pH range (4–8) common in soil. On the other hand, the accelerating effect of Fe oxide on the browning is extremely small. The promoting effects of Al and Si oxides were not observed in the conditions studied. The results obtained in the present study indicate that Mn(IV) oxides deserve close attention in the abiotic formation of humic substances in the environment.

* Permanent address: Faculty of Agriculture, University of Yamaguchi, Yamaguchi, Japan 753.

Table 3 ^3H -content of fracture fluids resulting from *in situ* ^3H -production, for various Li and B contents in the Carnmenellis granite*

Li-content (μg per g)	^3H -content†, TU for B-contents (μg per g)				
	15	50	250	500	1,000
50	0.14	0.11	0.05	0.03	0.02
250	0.57	0.47	0.24	0.15	0.08
500	0.95	0.81	0.45	0.29	0.17
1,000	1.40	1.25	0.76	0.52	0.31
1,500	1.67	1.52	1.00	0.71	0.44

* U = 13.6 p.p.m.; Th = 16.9 p.p.m. (ref. 15).

† For an equivalent total porosity of 1.0% and a rock bulk-density of 2.6 g cm^{-3} .

Table 1 Effects of Mn, Fe, Al and Si oxides on the browning of hydroquinone solution at different pH values at the end of the reaction period of 24 h

Oxide*	Initial pH of 4.0				Initial pH of 5.0				Initial pH of 6.0			
	Final		Absorbance		Final		Absorbance		Final		Absorbance	
	pH	Eh(V)	400 nm	600 nm	pH	Eh(V)	400 nm	600 nm	pH	Eh(V)	400 nm	600 nm
None	4.0	0.38	Colourless†		5.0	0.32	Colourless†		6.0	0.27	Colourless†	
Mn	4.1	0.44	1.64	0.46	5.3	0.38	0.88	0.20	7.5	0.32	7.44	1.84
Fe	4.0	0.40	0.06	0.02	5.0	0.33	0.01	Colourless†	6.1	0.27	Colourless†	
Al	4.0	0.38	Colourless†		5.0	0.32	Colourless†		6.0	0.27	Colourless†	
Si	4.0	0.38	Colourless†		5.0	0.32	Colourless†		6.0	0.27	Colourless†	

Oxide	Initial pH of 7.0				Initial pH of 7.8			
	Final		Absorbance		Final		Absorbance	
	pH	Eh(V)	400 nm	600 nm	pH	Eh(V)	400 nm	600 nm
None	6.9	0.24	0.30	0.09	6.9	0.24	0.70	0.17
Mn	7.7	0.29	8.62	2.06	7.7	0.32	8.86	2.12
Fe	6.9	0.25	0.10	0.02	7.0	0.25	0.37	0.10
Al	6.7	0.25	0.10	0.04	6.8	0.24	0.24	0.08
Si	6.8	0.24	0.20	0.06	6.9	0.24	0.49	0.12

* The experimental procedures were as stated in Fig. 1.

† The supernatant is visibly colourless and its absorbance is <0.01.

Noncrystalline or poorly crystalline oxides of manganese (birnessite, δ -MnO₂), iron, aluminium and silicon were prepared and used in this study as they commonly occur in soils¹⁶⁻¹⁹. Hydroquinone was used as the phenolic compound in this study, because benzoquinone derivatives are known to be parts of the chemical structure of humic acid^{2,3,5,6}, and the use of catechol derivatives which rapidly form coloured chelate complexes with iron²⁰ is inadequate for this study.

Figure 1 shows the absorption spectra of several browning solutions in the visible region at the end of the reaction period of 24 h. The colours of the solutions were light reddish brown to dark brown, and differed from the colour (yellow) of *p*-benzoquinone which is an oxidized monomeric product of hydroquinone. The shapes of their spectra varied with the reaction conditions, indicating that substances with varying degrees of polymerization were formed. Their absorbances in the visible region were distinct from each other. The degree of the browning of hydroquinone solution as influenced by the

various oxides can be compared by using light absorption at given wavelengths. The degree of humification of humic acid is commonly determined using light absorbances at 400 and 600 nm (ref. 9). Therefore, in this study, the absorbances at 400 and 600 nm were used to determine the effects of the oxides on the browning.

Table 1 shows the effects of Mn, Fe, Al, and Si oxides on the browning of hydroquinone solution at different pH values at the end of the reaction period of 24 h. The control solution without the oxides became coloured only when the initial pH of the solution was 7.0 or higher. The degree of the browning at pH 7.8 was higher than that at pH 7.0, indicating that oxidative polymerization was promoted at the higher pH. The addition of Mn oxide greatly accelerated the browning over the pH range (4.0-7.8). The degree of the browning in the Mn oxide system was influenced by the pH values. In the case of pH 5.0, the degree of the browning was much lower than that at other pH values. In both solutions of pH 4.0 and 5.0, the

Table 2 Effects of antiseptic and inoculum on the browning of hydroquinone solution both in the absence and presence of birnessite at the end of the reaction period of 24 h

No.	Mn oxide*	Antiseptic	Inoculum†	Final pH	Growth of micro-organisms‡	Absorbance	
						400 nm	600 nm
1	-	-	-	6.0		Colourless	
2	-	HgCl ₂	-	5.5		0.23	0.04
3	-	Toluene	-	6.0		Colourless	
4	-	-	+	6.0		Colourless	
5	+	-	-	7.4	ND§	6.70	1.66
6	+	HgCl ₂	-	7.1	ND	5.00	1.42
7	+	Toluene	-	7.4	ND	6.70	1.66
8	+	-	+	7.4		6.70	1.66

* 20 mg of Mn oxide (birnessite)²¹ was suspended in 20 ml of 0.1 M sodium acetate solution at pH 6.0 containing 10 μ mol ml⁻¹ hydroquinone; in certain cases, 2 ml of water, 2 ml of 1% HgCl₂, 1 ml of toluene or 2 ml of inoculum was added. The subsequent procedures are as stated in Fig. 1.

† Filtrate was obtained from the soil water suspension (soil:water ratio = 1:10). The fresh soil sample was taken from the Ap horizon of the Unity Orthic soil in Wilkie, Saskatchewan, Canada.

‡ The growth of microorganisms in the solution at the end of the reaction period of 24 h was examined using agar plates which contained nutrient broth and agar²⁴.

§ ND, no colonies were observed on the plates even after 4 days.

|| The supernatant is visibly colourless and its absorbance is <0.01.

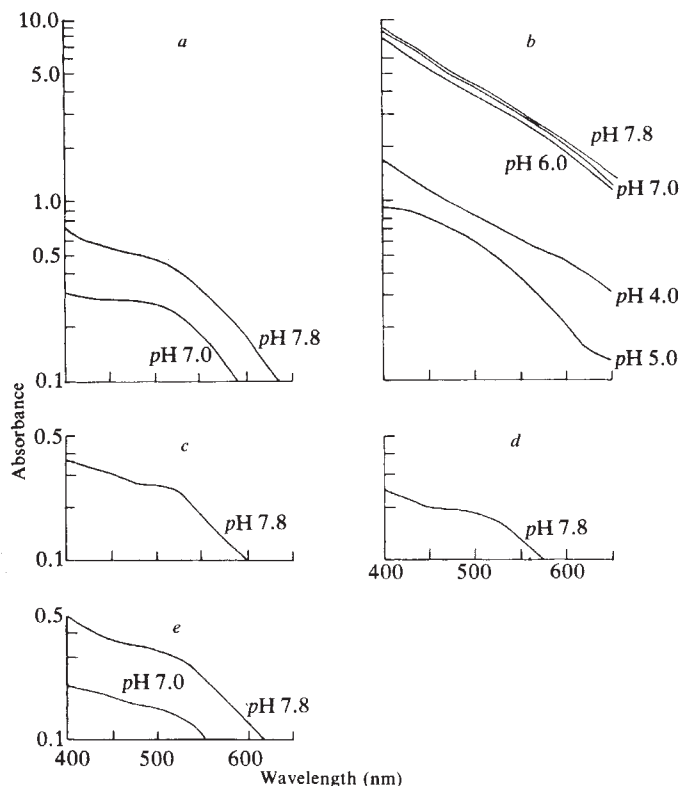


Fig. 1 The absorption spectra of several browning solutions in the visible region at the end of the reaction period of 24 h. *a*, Control (hydroquinone); *b*, Mn system; *c*, Fe system; *d*, Al system; *e*, Si system. pH means initial pH of the hydroquinone solution used. In systems *b* to *e* the Mn oxide (birnessite)²¹, Fe (ref. 22), Al (ref. 22), and Si (ref. 23) oxides were prepared. Twenty mg of each oxide was suspended in 20 ml of 0.1 M sodium acetate solution at different pH values containing 10 $\mu\text{mol ml}^{-1}$ hydroquinone. The flasks (125 ml) containing the suspensions were placed in an oscillating shaker in a water bath at 25 °C for 24 h. After the reaction period, the suspension was centrifuged at 38,000g for 20 min. The absorbance of the supernatant was measured with a spectrophotometer directly or after dilution; the resulting absorbances were scaled by the dilution factor. The sodium acetate solution which did not contain hydroquinone was used as reference.

Mn oxide added was dissolved at the end of the reaction period of 24 h. In the solution of pH 5.0, a very broad absorption band ranging from 400 to 600 nm was observed; this is different from the other Mn systems (Fig. 1*b*). Therefore, the low degree of the browning at pH 5.0 may be attributed to the formation of a stable complex which does not polymerize further. In the cases of pH 6.0 or higher, the degree of the browning was much higher than that at pH 4.0 or 5.0. This result cannot be explained by the *Eh* values of the system, because *Eh* decreased with increasing pH. In the Mn oxide systems to which were added solutions of initial pH values of 6.0, 7.0 and 7.8, the final pH values ranged from 7.5 to 7.7. On the other hand, final pH values in the control systems ranged from 6.0 to 6.9. Therefore, the high degree of the browning in the Mn oxide systems at the initial pH values of 6.0 or higher is attributed to the following processes: (1) the Mn oxide acts as a Lewis acid which accepts electrons from hydroquinone leading to its oxidative polymerization and (2) the acceleration of oxidative polymerization which is driven by an increase of pH which is in turn caused by the reduction of Mn oxide.

The absorption spectra of the browning solutions in the Mn systems at pH 4.0, 6.0, 7.0 and 7.8 (Fig. 1*b*) were similar to those of humic acids⁹. Their $\Delta\log K$, namely, the logarithm of the ratio of the absorbance at 400 nm to that at 600 nm (ref.

9), ranged from 0.55 to 0.62. This indicates that Mn oxides are capable of accelerating the conversion of phenolic compounds such as hydroquinone to humic acid-like polymers with a relatively high degree of humification.

The promoting effect of Fe oxide on the browning was observed only in acidic conditions, and the effect was higher at pH 4.0 than at pH 5.0 (Table 1). The accelerating effects of Al and Si oxides on the browning were not observed at all the pH conditions studied.

The control solution became coloured when the solution was adjusted to pH 7.0 or 7.8. In the presence of Fe, Al and Si oxides, a decrease in the degree of the browning was observed. This decrease is attributed to the adsorption of the browning substances by the oxides.

To determine whether microorganisms influence the browning of hydroquinone solution in the conditions studied, antiseptic or inoculum was added to the system. The browning of hydroquinone solution was then studied. As shown in Table 2, the hydroquinone solution in the absence of Mn oxide was colourless both in the absence and presence of toluene or inoculum. The degree of the browning in the Mn oxide system was not influenced by the presence of either toluene or inoculum. HgCl_2 slightly accelerated the browning of hydroquinone solution without Mn oxide, indicating that HgCl_2 has a weak power in causing the browning. The presence of Hg(II) in the Mn oxide system slightly suppressed the browning. HgCl_2 is a considerably weaker oxidant than MnO_2 (Table 2). Therefore, the influence of HgCl_2 on the browning of hydroquinone solution by MnO_2 may be attributed to the blocking of a portion of Lewis acid sites on the surfaces of MnO_2 by Hg(II) . Agar plates, inoculated with the browning solution, did not show any growth of microorganisms (Table 2). Therefore, the results clearly show that the browning of hydroquinone solution in the conditions studied occurs abiotically.

Our data reveal that Mn(IV) oxide is a very effective oxidant compared with Fe, Al and Si oxides in the abiotic browning of hydroquinone. This finding indicates that Mn(IV) oxide may have an important role in the formation of humic substances in the environment.

This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. We thank Dr J. J. Germida in our Institute for help with the microbiological aspects of the experiment.

Received 18 March; accepted 10 May 1982.

1. Kononova, M. M. *Soil Organic Matter* 2nd edn, 544 (Pergamon, Oxford, 1966).
2. Flaig, W., Beutelspacher, H. & Rietz, E. in *Soil Components* Vol. 1 (ed. Gieseking, J. E.) 1-211 (Springer, Berlin, 1975).
3. Felbeck, G. T. Jr *Adv. Agron.* **17**, 327-368 (1965).
4. Hurst, H. M. & Burges, N. A. in *Soil Biochemistry* Vol. 1 (eds McLaren, A. D. & Peterson, G. H.) 260-286 (Dekker, New York, 1967).
5. Martin, J. P. & Haider, K. *Soil Sci.* **111**, 54-63 (1971).
6. Schnitzer, M. & Khan, S. U. *Humic Substances in the Environment*, 327 (Dekker, New York, 1972).
7. Scheffer, F. & Ulrich, B. *Humus and Humusdüngung*, 260 (Ferdinand Enke, Stuttgart, 1960).
8. Kumada, K. & Kato, H. *Soil Sci. Pl. Nutr.* **16**, 195-200 (1970).
9. Kumada, K. *Chemistry of Soil Organic Matter* (in Japanese) 220 (University of Tokyo Press, 1977).
10. Kyuma, K. & Kawaguchi, K. *Proc. Soil Sci. Soc. Am.* **28**, 371-374 (1964).
11. Wang, T. S. C., Li, S. W. & Ferng, Y. L. *Soil Sci.* **126**, 15-21 (1978).
12. Wang, T. S. C., Kao, M. M. & Huang, P. M. *Soil Sci.* **129**, 333-338 (1980).
13. Haider, K., Frederick, L. R. & Flaig, W. *Plant and Soil* **22**, 49-64 (1965).
14. Scheffer, F., Meyer, B. & Niederbuddle, E. A. Z. *Pflanzenernähr. Düng. Bodenkd.* **87**, 26-44 (1959).
15. Haider, K., Martin, J. P. & Filip, Z. in *Soil Biochemistry*, Vol. 4 (eds Paul, E. A. & McLaren, A. D.) 195-244 (Dekker, New York, 1975).
16. Hsu, P. H. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) 99-143 (Soil Science Society of America, Madison, 1977).
17. Schwertmann, U. & Taylor, R. M. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) 145-180 (Soil Science Society of America, Madison, 1977).
18. McKenzie, R. M. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) 181-193 (Soil Science Society of America, Madison, 1977).
19. Wilding, L. P., Smek, N. E. & Dress, L. R. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) 471-552 (Soil Science Society of America, Madison, 1977).
20. Shindo, H. & Kuwatsuka, S. *Soil Sci. Plant Nutr.* **22**, 23-33 (1976).
21. McKenzie, R. M. *Miner. Mag.* **38**, 493-502 (1971).
22. Huang, P. M., Wang, T. S. C., Wang, M. K., Wu, M. H. & Hsu, N. W. *Soil Sci.* **123**, 213-219 (1977).
23. Mitchell, S. A. *Chem. Ind.* **23**, 924-933 (1966).
24. Germida, J. J. & Casida, L. E. Jr *J. Bact.* **144**, 1152-1158 (1980).

Is the CH₄, H₂ and CO venting from submarine hydrothermal systems produced by thermophilic bacteria?

John A. Baross, Marvin D. Lilley
& Louis I. Gordon

School of Oceanography, Oregon State University, Corvallis,
Oregon 97331, USA

Submarine hydrothermal vents are a major source of methane to the oceans^{1,2}. The methane, as well as H₂ and CO, are generally believed to result from degassing of the mantle or from abiogenic water-rock reactions¹, a conclusion supported by direct correlations between ³He and CH₄, and generally between CH₄, H₂ and CO and dissolved silicate in hydrothermal waters^{2,3}. An alternative source for these gases might be microbiological. This would imply that active bacterial communities exist in deep-sea hot water environments, some of which have temperatures exceeding 100 °C; this inference is without precedent. We have now found that the super-heated waters emanating from sulphide chimneys at 21°N along the East Pacific Rise and samples from the sulphide chimneys themselves harbour complex communities of bacteria capable of growing with generation times of 37–65 min, producing CH₄, CO, H₂ and traces of N₂O in media containing S₂O₃²⁻, Mn²⁺ and Fe²⁺ as energy sources, and oxidizing CH₄ at 100 ± 2 °C at 1 atm. These microbial communities consist of three to five morphologically distinct types and include both oxidative and anaerobic species. These mixed cultures will not grow at temperatures below 70–75 °C. Even though some of the communities originated from water of temperatures >300 °C, it is not known if they can grow and produce CH₄, CO and H₂ in super-heated waters kept liquid due to hydrostatic pressure. The discovery of these obligately thermophilic, gas-producing and consuming bacterial communities associated with submarine volcanic environments has interesting and important implications for prokaryotic evolution, marine geochemistry, industrial microbiology and exobiology.

Submarine hydrothermal systems located along tectonic rifts and ridges off the Galapagos Islands and along the East Pacific Rise^{4–6} and the Gorda Ridge have been studied and sampled since 1977. Two groups of vents have been described: (1) cracks and small fissures which emit warm water (5–22 °C) at a flow rate of ~2 cm s⁻¹ and (2) spectacular 3–10-m high conical sulphide mounds (chimneys) out of which spout super-heated waters with temperatures as high as 380 ± 30 °C and flow rates of metres per second (ref. 5). At the depth of these hydrothermal systems (2,500–2,650 m) seawater can remain liquid at temperatures up to ~460 °C due to the effect of hydrostatic pressure⁷. Sulphide chimneys are found at 21°N along the East Pacific Rise, and the warm water vents are found at both the Galapagos Rift and at 21°N. Associated with each of these hydrothermal environments are extensive and diverse groups of animals which are now known to derive their energy from chemosynthetic bacteria^{8,9}. These bacteria, which are found in large numbers in warm vent waters^{4,9}, on animal and rock surfaces^{10,11}, and as endosymbionts associated with specific vent animals¹², utilize S, metals and possibly organic gases and H₂ as energy sources. Bacteria had previously not been thought to exist in the super-heated waters associated with sulphide chimneys.

During October and November 1979, samples of hot vent water, sulphide chimney sections and associated fauna were obtained from the East Pacific Rise at 21°N using the manned submersible, *Alvin*. The present report is based on microbial cultures obtained from eight hot water samples from black and white smoking chimneys and one chimney wall sample. Initial chemical, physical and microbiological analyses of these waters were made on board the R/V *Gillis* shortly after the samples surfaced.

In general, the concentrations of the biologically important chemical constituents, H₂S, Mn and Fe, were high in all samples. The concentration of each chemical species was dependent on the degree of mixing with ambient seawater (Table 1). Degree of mixing is indicated by the difference in the water temperatures, both as measured with a temperature probe attached to the *Alvin* and as calculated from Mg numbers (see corrected temperature measurements, Table 1). Also, the presence of anaerobic and aerobic bacteria capable of growing at 25 °C in some of the samples indicate mixing of the sample with bacteria from ambient or warm vent water. This is particularly apparent

Table 1 Chemical and microbiological characteristics of hot water venting from sulphide chimneys at 21°N along the East Pacific Rise

Sample designation	Temperature (°C)		pH	Concentration of biologically important constituents (μM)				Direct counts (ml ⁻¹)	Total nos of bacteria		Presence of thermophiles
	Probe	Corrected		H ₂ S	Mn	Fe	Si		Viable counts at 25 °C (ml ⁻¹)		
									Aerobic	Anaerobic	
978 5/6	220	195	5.12	7.5	1,055	153	11,700	1.4 × 10 ⁴	<1	80	+
978 7/8	120	22	6.04	600	120	20	1,300	7.4 × 10 ⁵	<1	310	+
980 7/8	84	46	6.27	475	68	—	2,839	3.5 × 10 ⁵	<1	<1	+
979 11/12	100	3	7.53	1.2	17	56	200	1.8 × 10 ⁵	<1	25	+
979 1/2	120	6	6.96	3.4	46	14	390	3.6 × 10 ⁵	5	40	+
980 3/4	170	160	4.91	1,300	286	206	9,600	2.5 × 10 ⁵	<1	<1	+
980 5/6	30–60	34	6.46	5	96	137	2,100	No data	No data	No data	No data
982 7/8	290	306	4.24	4,300	800	—	18,320	4.7 × 10 ⁵	<1	<1	+
978(Niskin)*	3	3	Sample not analysed chemically					8.6 × 10 ⁵	5.3 × 10 ³	<1	—

All water samples were obtained with 750 ml gold-plated stainless steel and brass (one valve) samplers (provided by Sandia Laboratories, New Mexico) using the submergence research vehicle *Alvin* in October 1979. Subsamples were taken into sterile 20–150-ml bottles. Corrected temperature measurements were calculated from the relationship: $T = 0.00667 (°C \mu\text{mol}^{-1}) (\text{Mg})$ where $\text{Mg} = \text{Mg}_{\text{ambient}} - \text{Mg}_{\text{measured}}$; T = sample temperature. Mg data were provided by J. Edmond. All chemical measurements and pH were from John Edmonds (unpublished). Total direct bacterial counts were measured by epifluorescent microscopy using the procedures of Zimmerman and Meyer-Reil²¹. Water samples were immediately fixed on shipboard with 2% glutaraldehyde (Tousimis, electron microscopy grade) and maintained at 1 °C until counted back in the laboratory; total aerobic viable bacteria were enumerated using a basal salt medium (MS) which contained: NaCl, 26 g; KCl, 0.8 g; MgCl₂·6H₂O, 5.6 g; MgSO₄·7H₂O, 7.6 g; CaCl₂, 0.1 g in 1 l twice distilled water; MS was supplemented with Na₂S₂O₃·5H₂O, 0.5%; FeSO₄·7H₂O, 0.02%; MnSO₄·H₂O, 0.08%; Na₂MoO₄·2H₂O, 0.001%; Na₂HPO₄, 0.9%; KH₂PO₄, 0.15%; NaNO₃, 0.1% trace element mixture according to Rippka, *et al.*²²; medium adjusted to pH 7.5. Total anaerobic viable bacteria were enumerated using MS supplemented with 0.25% sodium formate, 0.1% NaNO₃, trace element mixture and phosphate buffer (above) at pH 7.5 and incubated in anaerobic jars in an atmosphere of H₂ and CO₂ using BBL Gas Pax disposable gas generator envelopes. Bacto purified agar (Difco) was dialysed in twice distilled water to remove soluble organic compounds and 12 g l⁻¹ was used. Water samples were enumerated using the spread plate technique for dilutions of 0.1, 0.01 and 0.001 ml whereas 1-, 5- or 10-ml samples were concentrated using 0.22-μm Millipore type GS membrane filter. The filters were then placed on the surface of the agar plates. The presence of thermophiles was detected using enrichment medium containing MS supplemented as described above for aerobic plating medium. Duplicate enrichments were adjusted to pH 5 and 7.5. A dilute heterotrophic medium 1/10 strength Lib-X (ref. 23) was also used. Glass screw-cap culture tubes (1.8 × 12 cm) containing 10 ml of the enrichment media were inoculated with 1–2 ml of the water sample. The inoculated samples were then placed in a boiling water bath for 1 h to minimize contamination and then incubated at 86 °C on a generator in the engine room of the R/V *Gillis*. All of the inorganic media enrichments at both pH 5 and 7.5 showed positive growth within 4 days. No growth was detected in the heterotrophic medium after 10 days.

* Sterile Niskin bag sample used to collect water from sites located adjacent to vents.

Table 2 Rate of H₂, CH₄ and CO production by 21° N hot water and sulphide-chimney associated bacterial communities at 100 ± 2 °C

Sample	Inoculum size (per 30 ml)	Medium Type	pH	Gas production			
				nmol per ml per h	nmol per ml per 9 h	nmol per ml per 9 h	Presence of N ₂ O (nmol ml ⁻¹)
978 5/6 (water)	1.3 × 10 ⁶	A(S-NH ₄ ⁺)	7	102	42	10	0.15 at 36 h
			5	71	39	9	ND
		C(formate)	7	31	27	11	0.12 at 36 h
		D(no energy source)	7	0.3	0.3	0.6	ND
982 7/8 (water)	3.0 × 10 ⁵	A(killed cells)*		ND	ND	ND	ND
		A	7	54	22	14	Trace‡
		C	7	(110)†	(46)†	(22)†	Trace‡
		D	7	0.2	ND	ND	ND
3140 (sulphide chimney rock)	6.1 × 10 ⁵	A(killed cells)		ND	ND	ND	ND
		B(S-NO ₃ ⁻)	7	24	24	7.5	Trace†
			5	23	22	10	ND
		C	7	26	25	10	1.3 at 36 h
		D	7	0.3	ND	ND	ND
		A(killed cells)		ND	ND	ND	ND

Inoculum size was determined using epifluorescent microscopy. All media were made with MS (see Table 1 legend) supplemented with trace element mix (Table 1) and buffered with 1.0 g l⁻¹ HEPES (Sigma). Medium A contained 0.02% FeSO₄·7H₂O; 0.06% MnSO₄·H₂O; 0.002% Na₂MoO₄·2H₂O; 0.02% (NH₄)₂SO₄; 0.2% Na₂S₂O₃·5H₂O; 0.03% Na₂HPO₄; 0.05% NaHCO₃. Medium B is the same as medium A except 0.02% KNO₃ was used in place of (NH₄)₂SO₄ as a nitrogen source. Medium C contained 0.02% FeSO₄·7H₂O; 0.06% MnSO₄·H₂O; 0.002% Na₂MoO₄·2H₂O; 0.02% (NH₄)₂SO₄; 0.02% sodium formate; 0.03% Na₂HPO₄. Medium D contained MS plus HEPES buffer and 0.03% Na₂HPO₄ and was used as a non-energy source medium. This medium was inoculated as described for other reactions. Cultures were inoculated into 60-ml serum bottles containing 30 ml medium. The serum bottles were capped with Teflon caps. All media were equilibrated at 98–100 °C before inoculation and capping. No gases were detected in any of the uninoculated medium controls. All gases were detected by gas chromatography using a modification of the procedure described by Swinnerton and Linnenbom²⁴ with a custom-built gas chromatograph equipped with helium ionization, flame ionization and electron capture detectors²⁵. In all cases most of the H₂ and CH₄ was produced by these cultures within 3 h of inoculation. The levels of CO continued to increase over a period of 8–12 h. The rate of gas formation was calculated based on three measurements made at 45 min, 2 h and 3 h for H₂ and CH₄, and at 45 min, 3 h and 9 h for CO. ND, not detected.

* Inoculum was exposed to 5% neutralized formalin for 24 h at 15 °C. The cells were then centrifuged at 12,000g for 10 min in a Sorvall RC-2B, resuspended in medium A and diluted to 10⁶ cells per ml after counting by epifluorescent microscopy.

† In these samples only one gas sample was measured after 30 h incubation. Based on other experimental results (see above) these concentrations of gases were probably reached within 3 h.

‡ Trace levels of N₂O were <0.1 nmol ml⁻¹ but >0.025 nmol ml⁻¹.

in samples 979 11/12 and 979 1/2. These samples also showed the greatest degree of mixing based on the Mg temperature calculations. In all samples, except the ambient Niskin bag sample (978), bacteria capable of growing at 86 °C were detected by enrichment in media composed of inorganic energy sources at either pH 5 or 7, but not in a dilute heterotrophic medium. The hottest water sample (982 7/8), which came from a black smoker, had the highest concentration of H₂S, Mn and Si and the lowest pH. In contrast, the other hot sample (978 5/6), which originated from a white smoker, had considerably lower levels of H₂S. High concentrations of H₂, CH₄ and CO were detected in both these samples. Also, in samples 979 3/4 and 982 7/8 the concentrations of CH₄ and H₂ exceeded 20 and 200 µM, respectively (M.D.L. and L.I.G., unpublished). The mixed communities of thermophilic bacteria have been kept intact by incubating at 85 °C with the transfer of 50 ml of culture into 450 ml medium at 6–8-week intervals. The experimental results reported here used these cultured communities.

The mixed cultures of thermophilic bacteria from these hot waters grew slowly at 80–86 °C and there was no evidence of gas production. When the cultured bacterial communities were transferred to fresh medium and incubated at 100 ± 2 °C in a water bath using a high boiling commercial coolant, rapid growth and active gas ebullition were observed. Table 2 shows the specific gases synthesized by the cultures and their rate of formation at 100 ± 2 °C. The rapid production of CH₄ by these communities was unexpected because no precautions had been taken to maintain anaerobic conditions with the initial water samples or with subsequent cultures. The gases were formed at both pH 5 and 7 in inorganic media and in a formate medium. In contrast to the H₂ and CH₄, which were at detectable levels within minutes after the samples were prepared, CO and N₂O required 9–24 h of incubation to reach measurable concentrations. The sulphide chimney sample also harboured a thermophilic microflora capable of producing the same suite of gases as the hot water cultures. This suggests either that there is a resident population of these organisms associated with sulphide chimneys which could be washed off into the venting fluid, or that these sites are contaminated with hot water bacteria. The temperatures of the outer surfaces of the chimneys

are not known, but observable schlieren indicate that they are considerably warmer than the ambient seawater (J.A.B. and L.I.G., unpublished observation).

The thermophilic communities investigated are composed of anaerobic and aerobic bacteria including methanogens and methylotrophs. Table 3 shows the rate of CH₄ oxidation by these cultures. Neither the proportion of methane oxidizers in the cultures nor the versatility of the organisms in oxidizing other gases and ammonia have been determined. Enrichments in CH₄ and methanol-containing media with the intent of isolating pure cultures have resulted in a predominance of bacteria having homogeneous morphological characteristics. However, due to the mixotrophic nature of the cultures (oxidize CH₄, Mn²⁺ and NH₄⁺), we are not yet positive of their purity.

The generation times of the mixed cultures at 100 °C in a medium containing S, Mn, Fe and NaHCO₃ ranged from 37 to 65 min (Table 3). The bacteria associated with the hottest water sample (982 7/8 at 304 °C) also had the shortest generation time. The lack of any measurable lag phase in the growth of these cultures at 100 °C might help explain their rapid production of CH₄ and H₂ (Table 2). Within 12–15 h incubation, the total numbers of bacteria remained constant at ~10⁹ cells ml⁻¹, although morphologically distinct groups continued to replace other groups for up to 2 weeks. Rod-shaped bacteria (1–2 µm) and small cocci (0.5–0.7 µm) dominated young cultures (up to 7 h) whereas in older cultures, sheathed, budding and filamentous bacteria, and a coccus having spines, were observed.

Because very few microorganisms have been described which can grow above 90 °C and only one species has been shown^{13–16} to grow above 100 °C, it is probable the vent thermophiles are new species. For example, the highest growth temperature previously reported for a methanogen and a methylotroph are ~70 and 55 °C, respectively^{14,17}. Also, because of the uniqueness of these marine thermophiles, nothing is known about the biochemical reactions involved in their formation of H₂, CH₄ and CO. The elucidation of these metabolic pathways awaits the isolation of pure cultures. Due to the growth temperature characteristics of these communities and their propensity to form clumps and sheaths, it has been difficult to obtain pure cultures.

Table 3 Growth rates and rates of methane oxidation by 21 °N thermophilic bacterial communities at 100 ± 2 °C

Sample	Growth rate		Generation time (min)	Methane oxidation rate	
	Time (h)	No. of bacteria (ml ⁻¹ × 10 ⁴)		¹⁴ C-CH ₄ oxidized to CO ₂ (nmol h ⁻¹)	Inoculum size (no. bacteria per 50 ml)
978 5/6 (water)	0	1.2	55	2.3	3.0 × 10 ⁷
	2	4.0			
	3.5	34			
	7	98			
	11	3,900			
928 7/8 (water)	0	2.7	37	2.5–3.9	5.0 × 10 ⁷
	2	15			
	3.5	150			
	7	2,800			
	11	11,000			
3140 (chimney rock)	0	2.4	65	3.1–3.9	1.5 × 10 ⁷
	2	10			
	3.5	25			
	7	120			
	11	3,700			

Growth rates of bacterial communities at 100 °C were determined using epifluorescent microscopy. The growth medium was MS with 0.5% HEPES buffer supplemented with 0.2% Na₂S₂O₃ · 5H₂O; 0.02% MnSO₄ · H₂O; 0.02% FeSO₄ · 7H₂O; 0.05% (NH₄)₂SO₄; 0.01 Ca(NO₃)₂; 0.05% Na₂HPO₄; 0.05% NaHCO₃; trace element mix (see Table 1 legend); 0.001% Bacto-yeast extract (Difco) and adjusted to pH 7.0 ± 0.2. Yeast extract to 0.001% was found to stimulate growth of the communities. Higher concentrations were inhibitory. The component(s) in yeast extract responsible for this growth response is not known. Growth rates were determined from duplicate 100-ml cultures in 150-ml screw-cap bottles. 10-ml subsamples from each culture were used in all the bacterial count measurements. Methane oxidation rates were measured using ¹⁴C-CH₄ at a final concentration of 1 μCi ml⁻¹ (50 mCi mmol⁻¹). The bacterial cultures were grown in MS supplemented with trace element mix, 0.02% NaNO₃; 0.03% Na₂HPO₄; 0.02% MnSO₄ · H₂O; 0.5% HEPES buffer in 60-ml serum bottles containing 25 ml medium with a gas mixture consisting of 20% methane and 80% air. After 24 h incubation at 100 °C, the cultures were brought up to a volume of 50 ml with MS medium, flushed for 2 min with air, capped and 1 ml of ¹⁴C-CH₄ was added. Three sets of triplicate samples were run for each culture and the rates of methane oxidation were measured after 1, 2 and 4 h. Oxidation of methane was terminated by adding 1 ml of 10 M NaOH. The samples were then transferred to 100-ml serum bottles fitted with rubber caps according to the procedures of Harrison *et al.*²⁶ and the ¹⁴C-CO₂ was collected on β-phenylethylamine-impregnated filter paper. The level of ¹⁴C-CH₄ incorporated into cells was determined as previously described²⁷. All samples were counted using a Beckman LS 100-C scintillation counter in 10 ml toluene fluor (Omnifluor, NEN). All counts were corrected for counting efficiency by the channel-ratio method. In general, about 1–5% of the total ¹⁴C-CH₄ taken up by these communities was incorporated into cellular material.

One of the important questions regarding the oxidative microbial activity in anaerobic submarine hydrothermal waters is the nature, source and concentrations of the electron acceptors. Oxygen and NO₃⁻, the most abundant electron acceptors in ambient seawater, were either not detected or were present in low concentrations in waters from the Galapagos vents⁴ or from 21° N chimneys (J. Edmond, personal communication). However, due either to the leaky nature of hydrothermal systems or to advection processes, it is possible that O₂ and NO₃⁻ are involved. As these electron acceptors would be the factors which most limit bacterial productivity, they would be rapidly utilized *in situ*. Many of the sulphur- and manganese-oxidizing bacteria isolated from the Galapagos and 21° N waters and animals are capable of coupling oxidation reactions with denitrification (J.A.B. unpublished data). An alternative hypothesis would be that other electron acceptors such as MnO₂ and N₂O could be abiogenically synthesized in hydrothermal waters.

Even though the temperatures of some water samples harbouring thermophiles were above 300 °C, we cannot conclude that these bacteria grow in those super-heated waters. As previously stated, seawater at the depths of these environments can remain liquid at temperatures up to 460 °C (ref. 7). Also, in some cases increased hydrostatic pressure can affect microorganisms in a manner analogous to a decrease in temperature¹⁸. Further experiments are needed to test the possibility that these thermophiles can grow in super-heated water at pressures equivalent to water depths of 2,500–3,000 m. Determining the maximum and optimum growth temperatures under such *in situ* pressures will help to characterize the distribution of actively growing thermophiles in submarine hydrothermal environments.

Although we do not know the extent to which microorganisms are contributing to the quantity of CH₄, H₂ and CO venting from hydrothermal systems, our preliminary findings indicate that these gases may not be produced entirely by abiogenic reactions. Particularly in the lower temperature vents, microbial formation of these gases may be very significant, but even in the high temperature vents at 21° N, this mechanism cannot be ruled out.

Finally, it has been suggested that the first group of microorganisms to evolve in the warm Archaean oceans were thermophiles and that submarine hydrothermal environments could be suitable sites for chemical and microbiological evolution to occur^{19,20}. It is tempting, therefore, to suggest that some of these present day marine hydrothermal vent thermophiles may represent physiological analogues to Archaean ocean bacteria.

Support for this research was provided by Office of Naval Research grant N00014-79C-0004 and NSF grant 79-27283. We also thank the crews of the R/V *Lulu* and R/V *Gillis* for their assistance, John Edmond for his invitation to participate on the expedition and for use of unpublished data, R. Y. Morita for use of laboratory facilities, and Clifford Dahm and Sarah Hoffman for their helpful discussions and editorial comments.

Received 15 April; accepted 24 May 1982.

- Welhan, J. A. & Craig, H. *Geophys. Res. Lett.* **6**, 829–831 (1979).
- Lupton, J. E. & Craig, H. *Science* **214**, 13–18 (1981).
- Lilley, M. D. & Gordon, L. I. *EOS* **60**, 863 (1979).
- Corliss, J. B. *et al. Science* **203**, 1073–1083 (1979).
- Spies, F. N. *et al. Science* **207**, 1421–1432 (1980).
- Hekinian, R., Fevrier, M., Bischoff, J. L., Picot, P. & Shanks, W. C. *Science* **207**, 1433–1444 (1980).
- Chen, C.-T. A. *Science* **211**, 298 (1981).
- Williams, P. M., Smith, K. L., Druffel, E. M. & Linick, T. W. *Nature* **292**, 448–449 (1981).
- Karl, D. M., Wirsén, C. O. & Jannasch, H. W. *Science* **207**, 1345–1347 (1980).
- Jannasch, H. W. & Wirsén, C. O. *Appl. envir. Microbiol.* **41**, 528–538 (1981).
- Hiatt, B. *Mosaic* **11**, 15–21 (1980).
- Felbeck, H., Childress, J. J. & Somero, G. N. *Nature* **292**, 291–293 (1981).
- Brock, T. D. *Thermophilic Microorganisms and Life at High Temperatures* (Springer, New York, 1978).
- Tansey, M. R. & Brock, T. D. in *Microbial Life in Extreme Environments* (ed. Kushner, D. J.) 159–216 (Academic, London, 1978).
- Zeikus, J. G. *Enzyme Microb. Technol.* **1**, 243–252 (1979).
- Heinen, W. & Lauwers, A. M. *Archs Microbiol.* **129**, 127–128 (1981).
- Zeikus, J. G., Ben-Bassat, A. & Hegg, P. J. *Bact.* **143**, 432–440 (1980).
- Baross, J. A., Hanus, F. J. & Morita, R. Y. *Appl. Microbiol.* **30**, 309–318 (1975).
- Fyfe, W. S. *Science* **213**, 105–110 (1981).
- Corliss, J. B., Baross, J. A. & Hoffman, S. E. *Oceanologica Acta* No. SP, 56–69 (1981).
- Zimmerman, R. & Meyer-Reil, L. A. *Kieler Meeresforschungen* **30**, 24–27 (1974).
- Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. *J. gen. Microbiol.* **111**, 1–61 (1979).
- Baross, J. A., Hanus, F. J. & Morita, R. Y. in *Effects of the Ocean Environment on Microbial Activities* (eds Colwell, R. R. & Morita, R. Y.) 180–202 (University Park Press, Baltimore, 1974).
- Swinerton, J. W. & Linnenbom, N. J. *J. chromat. Sci.* **5**, 570–573 (1967).
- Lilley, M. D., Baross, J. A. & Gordon, L. I. *Deep-Sea Res.* (in the press).
- Harrison, M. J., Wright, R. T. & Morita, R. Y. *Appl. Microbiol.* **21**, 698–702 (1971).
- Baross, J. A., Hanus, F. J., Griffiths, R. P. & Morita, R. Y. *J. Fish. Res. Bd Can.* **32**, 1876–1879 (1975).

Evidence that patterning mechanisms in developing and regenerating limbs are the same

Ken Muneoka & Susan V. Bryant

Developmental Biology Center and Department of Developmental and Cell Biology, University of California, Irvine, California 92717, USA

Some amphibians have the ability to form new limbs throughout their lives. The essential similarity between limb regeneration and the original development of the limb is that both involve the elaboration of new patterns of structures. While some investigators believe that the two developing systems use similar mechanisms to generate the limb pattern¹⁻⁷, others have stressed the basic differences in developing and regenerating limbs, and have concluded that different mechanisms exist⁸. Both developing and regenerating urodele limbs can be caused to form supernumerary parts following tissue misalignment so that cells from normally disparate positions within the limb circumference confront each other (developing limbs^{2,9,10}; regenerating limbs, see ref. 11 for review). If similar mechanisms are operating, the developing limb should respond to grafts of regenerating limb tissue, and vice versa, by forming supernumerary outgrowths. If different mechanisms are operating, either the grafts should fail or the response should be disorganized. Our results with the axolotl *Ambystoma mexicanum* suggest that the patterning mechanisms are the same.

Grafts between regenerating fore- and hindlimbs are as effective as grafts within the homologous limb in eliciting supernumerary limb formation^{12,13}. Similar conclusions can be drawn from experiments performed on the developing leg and wing buds of birds¹⁴⁻¹⁶. We exploited this in our experiments on the axolotl where development of the hindlimb is retarded relative to that of the forelimb¹⁷, thus enabling grafts to be made between animals of similar age in which the hindlimb is developing and the forelimb is regenerating. Forelimbs were amputated midway through the humerus at stage 46¹⁷ and allowed to regenerate to the stage of palette¹⁸. Stage 46 hindlimb buds were amputated near the base and exchanged with palette stage blastemas removed from older sibling larvae. A section through a developing hindlimb bud at stage 46 is shown in Fig. 1. The limb bud is undifferentiated at this stage. In the experimental series, contralateral grafts were made between developing and regenerating limbs so as to confront anterior and posterior tissue. In the controls, ipsilateral grafts were made between developing and regenerating limbs so as to align all circumferential positions. In addition, developing hindlimb buds were grafted either contralaterally (experimentals) or ipsilaterally (controls) to other hindlimb bud stumps (Fig. 2).

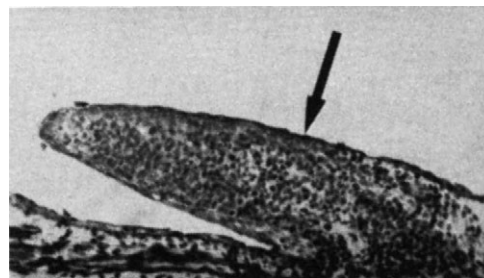


Fig. 1 A longitudinal section through a stage 46 hindlimb bud of the axolotl. No histological differentiation has yet occurred. The arrow indicates the level at which the limb bud was cut. Limbs were fixed in Bouin's and embedded in paraffin for serial sectioning. Sections were cut at 6 μ m and stained with Mallory's triple stain. $\times 70$.

The results are shown in Table 1. Control grafts, in which no positional disparity was created by grafting, mostly developed into normal limbs without supernumerary structures (Fig. 3a-c). In 3 out of 41 cases (one forelimb blastema to hindlimb bud stump and two hindlimb buds to forelimb blastema stumps), one extra digit formed at the anterior and/or posterior margin of the graft. The remaining 38 cases formed normal limbs according to the site of origin of the graft.

In the experimental series, where grafting created tissue disparities, supernumerary limbs formed in a high percentage of cases (Table 1). Contralateral exchange of developing hindlimb buds led to supernumerary limbs in 73% of the cases (Fig. 3f). Contralateral grafting of developing hindlimb buds to regenerating forelimb stumps led to supernumerary limbs in 70% of the cases (Fig. 3d), and the reciprocal experiment of contralateral grafting of regenerating forelimb blastema to developing hindlimb bud stump led to supernumerary limbs in 62% of the cases (Fig. 3e). In all the experimental series, extra structures formed either at the anterior, the posterior, or both borders of the graft junction (see Table 1), and in all cases the supernumerary limbs were of stump rather than graft handedness. Commonly, supernumerary limbs were not clearly separated from the grafted limb, and loss of digits along the line of symmetry between supernumerary and grafted limb was observed (see Fig. 3e). Similar types of limbs were observed by Iten and Bryant¹⁹ after contralateral grafts of early stage forelimb blastemas of newts (*Notophthalmus viridescens*) to appose anterior and posterior tissues. How such fusions can occur is discussed in terms of the polar coordinate model by Bryant *et al.*³.

The results of these experiments show that the interactions between developing and regenerating limbs of urodeles following confrontations between cells from disparate positions are similar to the interactions between similar limb bud and limb bud or blastema and blastema grafts. When developing axolotl forelimb buds were exchanged contralaterally to appose anterior and posterior tissues, Maden and Goodwin⁹ found that 70% of the cases produced limbs with supernumerary structures. Our results on the developing hindlimb bud are very

Table 1 Grafts between developing and regenerating limbs

Operation	Total	Normal*	Uninter- pretable	Supernumerary structures				Mean no. digits
				Total	Anterior	Posterior	Both	
Control series								
Hindlimb bud to hindlimb bud stump	12	12						
Hindlimb bud to forelimb blastema stump	17	15	2					
Forelimb blastema to hindlimb bud stump	12	11	1					
Experimental series								
Hindlimb bud to hindlimb bud stump	11	2 [†]	1 [‡]	8 (73%)	1	2	5	9.1
Hindlimb bud to forelimb blastema stump	20	2 [†]	4	14 (70%)	4	3	7	7.8
Forelimb blastema to hindlimb bud stump	13	2 [†]	3 [‡]	8§ (62%)	2	2	3	7.8

* Developing according to the site of graft origin.

[†] Limbs were of stump handedness.

[‡] Hypomorphic limbs.

[§] One case formed a ventral supernumerary structure.

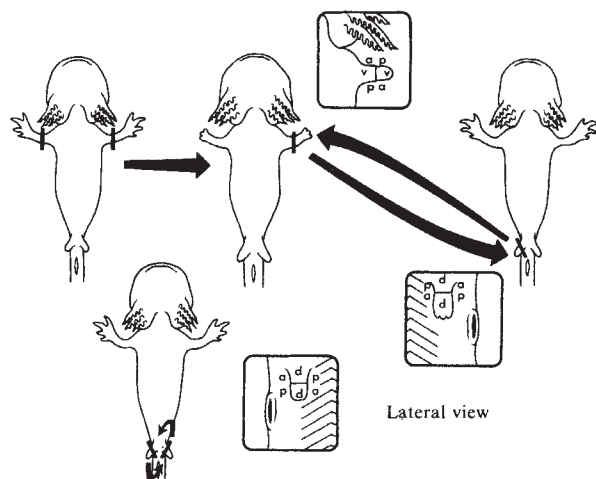


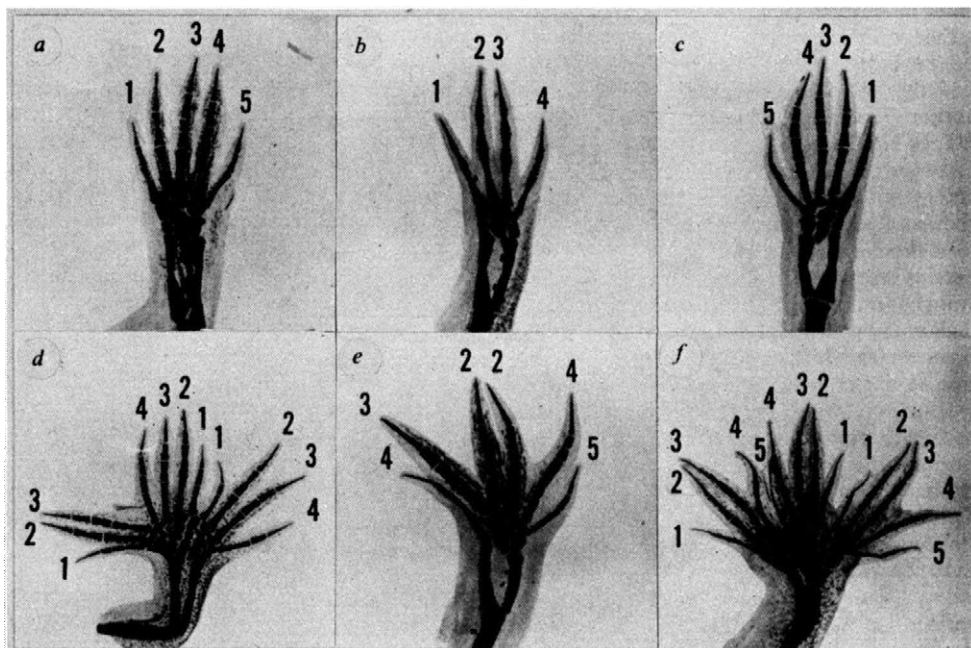
Fig. 2 Diagram of the grafting operations, shown here from the ventral aspect. *a*, The forelimbs of stage 46 (ref. 17) axolotls were amputated through the mid-humerus. Note that the forelimbs in the axolotl are fully differentiated when the hindlimbs are just beginning to grow out. After a period of 10–12 days the forelimb blastemas had regenerated to the stage of palette to early digits¹⁸. These animals were paired with more slowly developing stage 46 sibling axolotls. The regenerating forelimb blastema and the developing hindlimb bud were exchanged between the paired animals (usually between differently pigmented animals). Grafts were made either ipsilaterally (control series) matching anterior, posterior, dorsal and ventral limb tissues or contralaterally (experimental series) apposing anterior and posterior limb tissues. A contralateral graft is shown in the diagram. The insets show a closer ventral view of the limb bud to blastema stump graft and a lateral view of blastema to limb bud stump graft. *b*, The hindlimb buds of a stage 46 axolotl were exchanged either contralaterally (experimental series) as shown here, or simply amputated and replaced (control series). In the experimental series, the limb buds were oriented to appose anterior and posterior limb bud tissue while keeping dorsal and ventral limb bud tissue aligned. For all grafting operations animals were anaesthetized in MS222 (Eastman) diluted 1:3,000 in 50% Holtfreter solution. Operations were done in air with fine iridectomy scissors, forceps and microdissection scalpel. Operated animals were kept in individual dishes (100×25 mm plastic Petri dishes, Lab-tek), fed and changed daily, and checked (without anaesthetizing) every 2–3 days throughout the experiment. Camera lucida drawings were made every time the grafts were checked. In this way observations on the early stages of limb outgrowth were made and the positions of supernumerary outgrowths were recorded. *a*, Anterior; *d*, dorsal; *p*, posterior; *v*, ventral.

similar, with 73% of the cases producing supernumerary structures. Similar experiments by Thoms and Fallon¹⁰ on the developing forelimb buds of three species of urodeles (*A. mexicanum*, *Ambystoma maculatum* and *N. viridescens*) resulted in supernumerary structures in 95% of the cases. Experiments by Tank²⁰ on the regeneration blastema of *A. mexicanum* grafted to oppose anterior and posterior tissues showed that supernumerary structures were produced in 57% of the cases. Iten and Bryant¹⁹ showed in similar experiments in the newt that supernumerary structures were formed in 63% of the cases. In all these experiments on developing and regenerating limbs the position and handedness of the supernumerary structures which formed were similar, and are consistent with the polar coordinate model^{3,5}. These results, taken with those reported here, suggest that the developing and regenerating limbs of urodeles use similar mechanisms for pattern regulation during limb outgrowth.

It also appears from grafting of developing limb bud tissues from evolutionarily diverse vertebrate species into the chick wing bud, that common pattern regulatory mechanisms are shared among many different vertebrates^{14,21–23}. In these studies grafts of tissue from the posterior of the limb buds of mammals (man, hamster, mouse, pig and ferret), birds (Japanese quail, ringneck pheasant, chukar partridge and Peking duck) and reptiles (snapping turtle and painted turtle) into the anterior of chick wing buds resulted in the formation of supernumerary limb structures. Further, control grafts which were made so that no positional disparity was created resulted in the normal development of the host limb. The relationship between the developing amphibian limb bud and the developing chick limb bud is less clear because differences in the temperature tolerances of the two species make grafting experiments between them difficult to interpret²³. However, grafting studies on the amphibian limb bud show that its behaviour after apposition of normally nonadjacent tissues is remarkably similar to that of the chick limb bud^{2,9,10,24}. It seems therefore that the mechanisms of pattern formation in the development of vertebrate limbs are common to all vertebrates²². Our results here suggest that those patterning mechanisms common to vertebrate limb development are also the mechanisms that control patterning during limb regeneration in amphibians.

This conclusion may have important implications for the way we think about how the vertebrate limb pattern is formed as

Fig. 3 Skeletal preparations of control and experimental limbs, prepared by the Victoria blue technique²⁵ 4 to 6 weeks after the grafting operation. *a*, A control ipsilateral graft of a developing right hindlimb bud put onto a regenerating right forelimb stump. A normal five-digit right hindlimb was formed on the forelimb stump. *b*, A control ipsilateral graft of a right regenerating forelimb blastema put onto a developing right hindlimb bud stump. The graft produced a normal four-digit right forelimb on the hindlimb stump. *c*, A control ipsilateral graft of a developing left hindlimb bud put onto a developing left hindlimb bud stump. The resulting limb is a normal five-digit left hindlimb. *d*, A limb which resulted from the contralateral graft of a left developing hindlimb bud to a right regenerating forelimb blastema stump. Supernumerary limb structures are present both anterior and posterior to the grafted hindlimb. *e*, A limb which resulted from the contralateral graft of a left regenerating forelimb blastema onto a right developing hindlimb bud stump. Digits derived from the graft are located on the left of the figure; supernumerary limb structures are present on the right. The two limbs have fused during outgrowth with loss of digits along the line of symmetry between them. *f*, A limb resulting from a contralateral graft of a left developing hindlimb bud to a right developing hindlimb bud stump. Supernumerary limb structures are present both anterior and posterior to the grafted hindlimb. *a–f*, ×5.



limb development and limb regeneration have mostly been regarded as two very different fields of study. Models of pattern formation during limb development have been drawn solely from data generated from work on limb development and the same is true for most models of pattern formation during limb regeneration. As there is now direct evidence that the mechanisms controlling pattern formation during limb development are the same as those for limb regeneration, we believe that the experimental data from both limb development and limb regeneration studies must be considered together to understand pattern formation. Why all vertebrates can develop limbs, but only a few can regenerate them remains puzzling. However, we propose that during evolution, the underlying patterning mechanisms have remained the same, but that the ability to

call these mechanisms into use has been modified. In other words, the urodele limb can regenerate because it has maintained its ability to implement pattern mechanisms whereas the chick wing cannot regenerate because it loses this ability very early in its development.

These studies therefore show that it is possible to learn about limb development by studying limb regeneration. Conversely, studies of limb development may help us to understand the causes of failure of limb regeneration in higher vertebrates.

We thank Christine Rollman-Dinsmore, William P. Hayes, Laurie E. Iten, Nancy A. O'Rourke and L. David Wise for their comments and Warren F. Fox and Kate John-Alder for assistance. This research was supported by NIH grants HD 06082 and HD 07029.

Received 26 February; accepted 14 May 1982.

1. Bryant, S. V. & Iten, L. E. *Dev Biol.* **50**, 212-234 (1976).
2. Maden, M. *J. embryol. exp. Morph.* **63**, 243-265 (1981).
3. Bryant, S. V., French, V. & Bryant, P. J. *Science* **212**, 993-1002 (1981).
4. Carlson, B. M. in *XIX Morphological Congr. Symp.* 135-140 (Charles University, Prague, 1978).
5. French, V., Bryant, P. J. & Bryant, S. V. *Science* **193**, 969-981 (1976).
6. Faber, J. *Adv. Morph.* **9**, 127-147 (1971).
7. Stocum, D. L. *Differentiation* **3**, 167-182 (1975).
8. Tickle, C. *Am. Sci.* **69**, 639-646 (1981).
9. Maden, M. & Goodwin, B. C. *J. Embryol. exp. Morph.* **57**, 177-187 (1980).
10. Thoms, S. & Fallon, J. *J. Embryol. exp. Morph.* **60**, 33-55 (1980).
11. Tank, P. W. & Holder, N. Q. *Rev. Biol.* **56**, 113-142 (1981).
12. Rollman-Dinsmore, C. G. & Bryant, S. V. *J. exp. Zool.* (in the press).

13. Stocum, D. L. *Dev Biol.* **79**, 276-295 (1980).
14. Balcuns, A., Gasseling, M. T. & Saunders, J. W. *Jr Am. Zool.* **10**, 323 (1970).
15. Saunders, J. W. *Jr Ann. N.Y. Acad. Sci.* **193**, 29-42 (1972).
16. Saunders, J. W. Jr & Gasseling, M. T. in *Epithelial-Mesenchymal Interactions* (eds R. Fleischmajer & R. E. Billingham) 78-97 (Williams & Wilkins, Baltimore, 1968).
17. Schreckenberg, G. M. & Jacobson, A. G. *Dev Biol.* **42**, 391-400 (1975).
18. Tank, P. W., Carlson, B. M. & Connelly, T. G. *J. Morph.* **150**, 117-128 (1976).
19. Iten, L. E. & Bryant, S. V. *Dev Biol.* **44**, 119-147 (1975).
20. Tank, P. *Dev Biol.* **62**, 143-161 (1978).
21. McCabe, J. A. & Parker, B. W. *J. exp. Zool.* **195**, 311-317 (1976).
22. Tickle, C., Shellsell, G., Crawley, A. & Wolpert, L. *Nature* **259**, 396-397 (1976).
23. Fallon, J. F. & Crosby, G. M. in *Vertebrate Limb and Somite Development* (eds D. A. Ede, J. R. Hinchliffe & M. Balls) 55-69 (Cambridge University Press, 1977).
24. Cameron, J. & Fallon, J. F. *Dev Biol.* **55**, 320-330 (1977).
25. Bryant, S. V. & Iten, L. E. *Wilhelm Roux Archiv.* **174**, 90-101 (1974).

Selective collateral elimination in early postnatal development restricts cortical distribution of rat pyramidal tract neurones

Brent B. Stanfield, Dennis D. M. O'Leary & Carl Fricks

The Salk Institute for Biological Studies and The Clayton Foundation for Research—California Division, PO Box 85800, San Diego, California 92138, USA

The pyramidal tract, comprising those axons which pass from the neocortex to the medulla and spinal cord, is among the most thoroughly studied projections of the mammalian cortex. Recent studies using anterograde axon tracing techniques have provided information concerning the time course of the growth of pyramidal tract fibres^{1,2}, yet much remains to be learned about its development. We have now begun to study the distribution of the neurones of origin of the pyramidal tract during the postnatal development of the rat neocortex using the recently introduced retrogradely transported fluorescent marker, True blue³. During the first postnatal week, injections of True blue into the pyramidal decussation result in the labelling of pyramidal tract neurones which are distributed virtually throughout the tangential extent of layer V of the neocortex, whereas after comparable injections during the fourth postnatal week the distribution of such cells is much more restricted and remains restricted into adult life. This developmental restriction is most dramatic in the occipital cortex: during the first postnatal week many pyramidal tract neurones are found throughout the visual cortex while none is seen in this area in the adult. When True blue is injected into the pyramidal decussation during the first postnatal week and the animals are allowed to survive until the fourth postnatal week, the distribution of pyramidal tract neurones is as widespread as in the immediate postnatal period and includes the entire visual cortex. This implies that many of the neurones in the occipital cortex initially send a collateral into the pyramidal tract which is later eliminated, although the neurones themselves persist. These findings, together with similar recent observations on the development of the callosal connections^{4,5}, indicate that the elimination of axon collaterals may be a general feature of the development of cortical projection systems, and that such transitory collaterals may traverse considerable distances.

The experimental procedure in all instances involved the injection of 0.3–1.0 µl of the fluorescent dye (a 2–5% solution of True blue) into the pyramidal tract at the level of its decussation at the spino-medullary junction. This injection site is convenient because: (1) of the ease with which consistent injections can be made; (2) of the considerable distance between this site and the targets of non-pyramidal tract cortical neurones; (3) it is known that in the rat the fibres of the pyramidal tract have reached the decussation at the time of birth^{2,6}. The pyramidal decussation of the rat consists of corticospinal axons (which after crossing descend into the spinal cord) and crossed corticobulbar fibres which terminate in the dorsal column and other medullary nuclei^{7,8}. Thus, we will refer to the retrogradely labelled cortical neurones in our experiments as pyramidal tract

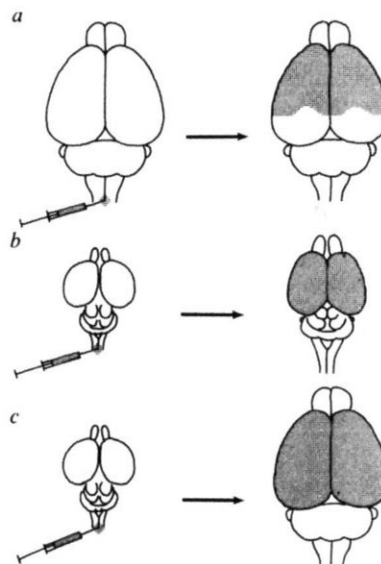


Fig. 1 Schematic illustrations of the three experimental paradigms used in this study, summarizing the results of each. *a*, When True blue is injected into the pyramidal decussation of adult rats, or rats at 20 days of age, the distribution of retrogradely labelled pyramidal tract neurones is widespread but limited to about the rostral two-thirds of the cerebral hemisphere. *b*, *c*, Following similar injections in 2-day-old rat pups, labelled cells are spread throughout the tangential extent of the cortex, whether the animals are killed on day 6 (*b*) or are allowed to survive until day 25 (*c*).

neurons without distinguishing between the corticobulbar and corticospinal projections because both of these had access to the injected label. After a suitable survival period the animals were perfused with 10% buffered neutral formalin, the brains were removed and processed as described elsewhere⁹ and the frozen sections were examined with a Leitz Dialux-20 fluorescence microscope.

Following injections of True blue (0.5–1.0 μ l) into the pyramidal decussation of adult rats ($n = 7$), the distribution of labelled cells in the cortex after a 5–7 day survival period is similar to, but more extensive than, that of corticospinal cells, as recently shown through the use of similar techniques^{10,11}. Although strictly limited in the radial dimension of the cortex to layer V (as identified in adjacent Nissl-stained sections), the labelled neurones occupy a large and continuous tangential sheet of cortex. This sheet of labelled cells spans virtually the entire frontal cortex and extends uninterruptedly into the parietal cortex (Fig. 1a). An essentially identical distribution of pyramidal tract neurones is seen in rats that are injected on postnatal day 20 and allowed to survive for 5 days ($n = 14$). We emphasize that although the extent of the cortex in which pyramidal tract neurones are found varies slightly from case to case (due no doubt to slight variations in the effective injection site), in none of these animals, nor in any of the adult cases, are retrogradely labelled neurones identifiable in the occipital cortex (Figs 2a, 3a).

In animals injected with True blue (0.3–0.5 μ l) on postnatal day 2 and killed on postnatal day 6 ($n = 17$), a large number of retrogradely labelled cells can be seen in layer V of the cortex (Fig. 3b). In fact, the density of labelled cells in the early postnatal cortex is appreciably greater than that in the older animals. The band of labelled cells is spread across essentially the entire tangential extent of the cortex, from the frontal cortex to the caudal pole of the hemisphere (Fig. 1b). We emphasize that in none of the present experiments did the injected dye spread beyond the lower medulla and that we are confident that there was no involvement of the corticopontine projection.

Thus, during the first week of postnatal life pyramidal tract neurones seem to be confined to layer V, but are distributed throughout the tangential extent of the neocortex, whereas by the fourth week such cells are excluded from the posterior pole

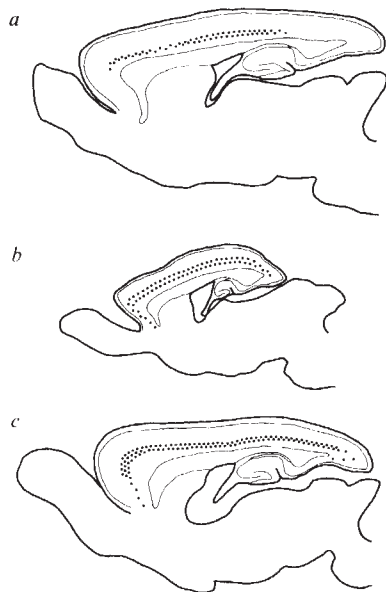


Fig. 2 Tracings of sagittal sections of rat brains at approximately the same medio-lateral position to show the distribution of retrogradely labelled neurones in the cortex of animals which: a, received an injection of True blue in the pyramidal decussation on postnatal day 20 and was killed on postnatal day 25; b, received a similar injection on postnatal day 2 and was killed on postnatal day 6; c, was injected on postnatal day 2 and killed on postnatal day 25. Each dot represents ~ 15 labelled neurones.

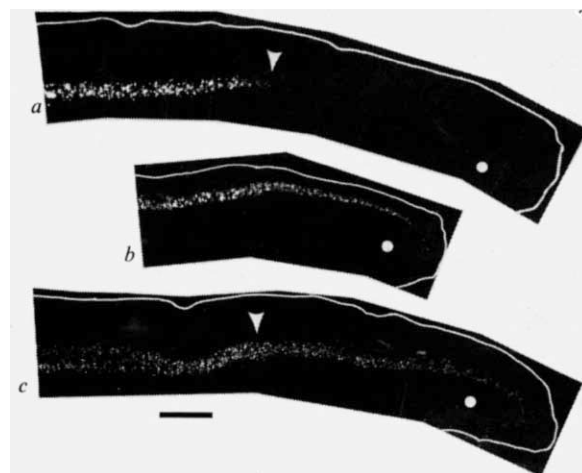


Fig. 3 These photomontages illustrate the distribution of True blue-labelled neurones in sagittal sections of the caudal part of the cortex from animals injected in the pyramidal decussation: a, at postnatal day 20 and surviving until postnatal day 25; b, at postnatal day 2 and surviving until postnatal day 6; c, at postnatal day 2 and surviving until postnatal day 25. The white line indicates the cortical surface; the splenium of the corpus callosum is marked with a large dot. The arrowhead in a marks the caudal limit of the band of labelled cells and that in c is at the same level. Scale bar, 1 mm.

of the hemisphere, and in particular from the visual cortex^{12,13}. This difference in the distribution of pyramidal tract neurones at these two ages is most clearly seen in sagittal sections as shown in Figs 2 and 3. To determine whether the developmental restriction in the distribution of pyramidal tract neurones is brought about by the selective elimination of axon collaterals or by cell death, in 15 animals we injected True blue (0.3–0.5 μ l) into the pyramidal decussation on postnatal day 2 and allowed the rats to survive until postnatal day 25. As we have previously reported⁵, True blue remains within labelled neurones for several weeks, without significant degradation or leakage. In these experiments the distribution of labelled cortical neurones is as widespread as that in animals which were injected on postnatal day 2 and allowed to survive for only a few days even though the adult distribution of pyramidal tract neurones is established by the time these animals were killed on day 25 (Fig. 1). Of course, the cortex in these animals had continued its normal course of maturation and was cytoarchitecturally indistinguishable from that of control animals of the same age. The band of retrogradely labelled neurones was again confined to layer V, but extended throughout the cortex, including the visual cortex, from which, as we have shown, pyramidal tract neurones are normally excluded by the fourth postnatal week (Figs 2, 3).

The most straightforward interpretation of these findings is that early in postnatal life there are neurones throughout the tangential extent of layer V which send axons into the pyramidal tract, and that during the next 3 weeks many of these cells (and especially those in the visual cortex) lose their pyramidal tract collaterals even though the neurones themselves survive. At present we cannot rule out the possibility that neuronal cell death also contributes to the exclusion of pyramidal tract neurones from the occipital cortex during postnatal development, but certainly this restriction is not due mainly to the death of these neurones, because when labelled by way of their pyramidal tract axon on postnatal day 2, many of these cells are still identifiable at 4 weeks of age. This result also excludes the possibility that the developmental change in the distribution of labelled neurones is due to a secondary, tangential migration of neurones. Although we cannot determine how many neurones lose pyramidal tract collaterals during postnatal development, within the occipital cortex collateral elimination is clearly a substantial event as it results in the disappearance of all pyramidal tract axons which initially arise from this part of the cortex.

These observations on the development of pyramidal tract neurones are similar to those that we and others have recently made on the development of callosally projecting neurones^{4,5}. In that instance, however, the developmental change results in a refinement of the distribution of callosal neurones within an area, whereas in the present study we have shown the complete elimination of a substantial projection from an entire region of cortex. However, together these studies suggest that a loss of collateral projections may be a rather general feature of the development of cortical efferent systems. We have not examined the time course of the restriction in the distribution of pyramidal tract neurones in detail, but the data available so far suggest that it is similar to that affecting the distribution of callosal neurones in the rat¹⁴. It would also be of interest to determine if some of early pyramidal tract collaterals reach and form synapses in the spinal cord before they are withdrawn. At present, it is not known whether collateral elimination of this type also involves synapse elimination—a phenomenon known to occur in several developing systems¹⁵. Interestingly in one of the systems in which synapse elimination has been shown to occur, there is apparently no change in the number of neurones which give rise to these synapses¹⁶. Finally, we emphasize that during the first postnatal week, when the tangential distribution of the pyramidal tract neurones is widespread, their radial distribution within layer V has already been determined.

We thank Madeline Farris and Kris Trulock for technical assistance, Pat Thomas for secretarial help, and Dr W. M. Cowan for support, advice and encouragement. This work was supported in part by NIH grants NS-16980, EY-03653 and GM-07198.

Received 1 March; accepted 14 May 1982.

1. Reh, T. & Kalil, K. J. *Comp. Neurol.* **200**, 55–67 (1981).
2. Schreyer, D. J. & Jones, E. G. *Neuroscience* (in the press).
3. Bentivoglio, M., Kuypers, H. G. J. M., Catsman-Berrevoets, C. E. & Dann, O. *Neurosci. Lett.* **12**, 235–240 (1979).
4. Innocenti, G. M. *Science* **212**, 824–827 (1981).
5. O'Leary, D. D. M., Stanfield, B. B. & Cowan, W. M. *Dev. Brain Res.* **1**, 607–617 (1981).
6. DeMyer, W. *Archs Neurol.* **16**, 203–211 (1967).
7. Valverde, F. J. *comp. Neurol.* **119**, 25–54 (1962).
8. Wise, S. P. & Jones, E. G. *J. comp. Neurol.* **175**, 129–158 (1977).
9. Sawchenko, P. E. & Swanson, L. W. *Brain Res.* **210**, 31–51 (1981).
10. Catsman-Berrevoets, C. E., Lemon, R. N., Verburgh, C. A., Bentivoglio, M. & Kuypers, H. G. J. M. *Expl Brain Res.* **39**, 433–440 (1980).
11. Catsman-Berrevoets, C. E. & Kuypers, H. G. J. M. *Brain Res.* **218**, 15–33 (1981).
12. Woolsey, C. N. in *Milbank Symp. on the Biology of Mental Health and Disease*, 193–206 (Hoebner, New York, 1952).
13. Montero, V. M. *Brain Res.* **53**, 192–196 (1973).
14. Ivy, G. O. & Killackey, H. P. *J. comp. Neurol.* **195**, 367–389 (1981).
15. Purves, D. & Lichtman, J. W. *Science* **210**, 153–157 (1980).
16. Johnson, D. A. & Purves, D. *J. Physiol., Lond.* **318**, 143–159 (1981).

Glycogenolysis induced by serotonin in brain: identification of a new class of receptor

T. T. Quach, C. Rose, A. M. Duchemin & J. C. Schwartz

Unité de Neurobiologie (U 109), Centre Paul Broca de l'INSERM, 2ter rue d'Alésia, 75014 Paris, France

Serotonin-containing neurones in brain have been proposed to have a role in the control of physiological mechanisms such as sleep, thermoregulation, pain perception and endocrine secretions as well as in the physiopathology of migraine or depressive illness^{1–6}. One difficulty in testing these possibilities lies in the scarcity of pharmacological agents able to interact selectively with the probably multiple classes of serotonin receptors in the central nervous system. Development of such agents would be facilitated by simple *in vitro* models in which biological responses to serotonin in mammalian brain could be quantified. Thus a serotonin-sensitive adenylate cyclase has been characterized in rat brain, but the response to serotonin is weak in newborn and practically absent in adult animals^{7–11}.

In addition, two pharmacologically distinct classes of serotonergic binding site have been identified using ³H-serotonin and ³H-spiperone as ligands^{12–17}, but their identification as receptors remains to be established. More recently, serotonin has been shown to stimulate phosphorylation of a neuronal protein in slices from the facial motor nucleus, although the receptors mediating this action were not characterized¹⁸. We now report that serotonin stimulates glycogen hydrolysis in slices of cerebral cortex, that this action is mediated by a novel class of receptors and that tricyclic antidepressants are among the best competitive antagonists of the indolamine.

The addition of serotonin to slices from mouse cerebral cortex preincubated with ³H-glucose for 30 min, at a time when ³H-glycogen level in tissue has reached a plateau¹⁹, results in rapid hydrolysis of the ³H-polymer. The serotonin-induced glycogenolysis is concentration dependent, occurs with a K_{act} of $20 \pm 3 \mu\text{M}$ and reduces the ³H-glycogen level by, at the most, $78 \pm 4\%$ (Fig. 1a). A similar effect is observed with other indolamine derivatives such as tryptamine or 5-methoxytryptamine, while tryptamine derivatives *N*-methylated on the side chain, such as *N*-methyltryptamine, bufotenine, psilocybin and psilocin, are only partially active and *N,N*-dimethyltryptamine is totally inactive (Fig. 1a, b). This latter observation prompted us to study *N,N*-dimethyltryptamine as antagonist: we found that increasing concentrations of this compound shift the concentration–response curve to serotonin rightwards, without modifying the maximal response (Fig. 2). This suggests that *N,N*-dimethyltryptamine acts as a competitive antagonist with

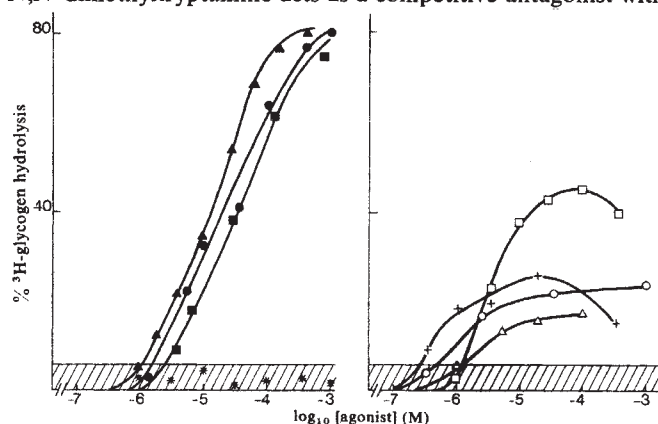


Fig. 1 ³H-glycogen hydrolysis induced by serotonin and other tryptamine derivatives in slices from cerebral cortex. Male Swiss albino mice (18–20 g, Lesieux, France) were killed by decapitation, the cerebral cortex was quickly removed and slices (250 × 250 μm thick) were prepared in a cold room (4 °C) with a McIlwain tissue slicer. Pooled cortical slices from six animals were preincubated for 15 min at 37 °C in a slightly modified Krebs–Ringer bicarbonate medium (120 mM NaCl, 5 mM KCl, 2.6 mM CaCl₂, 0.67 mM MgSO₄, 1.2 mM KH₂PO₄, 3 mM glucose and 27.5 mM NaHCO₃) under a constant stream of O₂/CO₂ (95:5). The slices were then washed once with fresh medium, 300- μl aliquots of the tissue suspension corresponding to 0.5 mg protein were distributed in incubation tubes and 13.4 nmol of ³H-glucose (500 mCi mmol⁻¹; Amersham) were added in a volume of 10 μl . Test agents were added when ³H-glycogen level in tissue had reached a steady state, that is, after a 30 min incubation¹⁹, and the slices were further incubated for 20 min. The incubations were stopped by rapid centrifugation. The supernatant was discarded and replaced by 300 μl of fresh medium in which the slices were sonicated. A fraction of the resulting homogenate was then immediately deproteinized by heating at 95 °C for 10 min and ³H-glycogen assayed by the ethanol filter paper technique¹⁹. Results are expressed as per cent of the basal ³H-glycogen level (26.3 ± 2.1 d.p.m. $\times 10^3$ per mg protein). The shaded area corresponds to the s.e.m. of basal ³H-glycogen level. Values are the mean of 15–25 determinations performed in 3–5 independent experiments. $\blacktriangle$, Serotonin; $\ast$, *N,N*-dimethyltryptamine; $\bullet$, 5-methoxytryptamine; $\blacksquare$, tryptamine; $\triangle$, *N*-methyltryptamine; $+$, bufotenine (5-hydroxy-*N,N*-dimethyltryptamine); $\circ$, psilocybin (*O*-phosphoryl-4-hydroxy-*N,N*-dimethyltryptamine); and $\square$, psilocin (4-hydroxy-*N,N*-dimethyltryptamine).

an apparent K_i of 0.8 μM . The ergot derivative (+)LSD can elicit complete ^3H -glycogen hydrolysis but only at high concentrations; its action was also antagonized by N,N -dimethyltryptamine with an apparent K_i in the micromolar range. Typical serotonergic antagonists such as methysergide, cyproheptadine and methiothepin fail to or weakly block the serotonin-induced response (Table 1).

Among neuroactive drugs, the antipsychotics chlorpromazine, spiperone, pipamperone and pimozide antagonize the serotonin-induced glycogenolysis with inhibition constants of between 4 and 20 μM ; various tricyclic antidepressant compounds competitively block this response, all with inhibition constants of $\sim 1 \mu\text{M}$.

The glycogenolytic response elicited by serotonin in cortical slices suggests that this amine should be added to the list of neurotransmitters controlling carbohydrate metabolism in brain and which comprises the other glycogenolytic agents noradrenaline¹⁹⁻²¹ (via β_1 receptors), histamine²² (via H_1 receptors) and the neuropeptide vasoactive intestinal peptide (VIP)²³. The serotonin-induced glycogenolysis requires higher concentrations of amine, however, particularly in comparison with VIP. Because the response to these agents occurs in the same cells (as shown by the almost total ^3H -glycogen hydrolysis in all cases), lower serotonin concentrations might be expected to be required in the presence of other glycogenolytic agents.

In a preparation such as brain slices, an effect observed with one neurotransmitter could be indirectly mediated by the release of another. The lack of antagonism by β -adrenergic and H_1 -receptor blockers (Table 1) indicates that neither noradrenaline nor histamine is involved. Experiments with phosphodiesterase inhibitors or with various concentrations of calcium indicate that phosphorylase activation leading to glycogen hydrolysis seems to result in brain²⁰⁻²², as in other tissues²⁴⁻²⁶, from one of two mechanisms: stimulation of an adenylate cyclase, as in the case of noradrenaline^{19,20} and VIP²³, or translocation of calcium ions, as in the case of histamine²². The former mechanism is probably not involved in the response

to serotonin as it is not modified in the presence of 0.7 μM isobutyl methylxanthine, a phosphodiesterase inhibitor, but is diminished when the calcium concentration is lowered (not shown). This suggests that VIP is probably not indirectly responsible for the serotonin effect, but this requires confirmation with antagonists of the neuropeptide, which are not available.

Serotonin induces phosphorylation of the 'protein I' in slices from facial motor nucleus at similar concentrations to those required for the glycogenolytic response, and both responses are blocked by 10 μM mianserin; however, only the former is potentiated by isobutyl methylxanthine¹⁸. Among the several other unidentified proteins which were phosphorylated in response to serotonin, one may correspond to phosphorylase.

Glycogen is the single largest energy reserve in the brain and the glycogenolytic effect of serotonin should result in an increased availability of glucose. The cellular location of this response remains to be identified, however, as the polymer is present both in neuronal and glial elements²⁷. Nevertheless the observations that impairment of monoaminergic neurotransmission in mice subchronically treated with reserpine results in the development of hypersensitivity to the glycogenolytic actions of noradrenaline²⁰ as well as to those of serotonin (T.T.Q., C.R., A.M.D. and J.C.S., in preparation) suggest that the target cells are tonically controlled by the two monoaminergic systems. Interestingly, the three aminergic systems controlling glycogenolysis—the coeruleo-cortical noradrenergic²⁸, the raphe-cortical²⁹ and the histaminergic ascending³⁰ projections—are highly divergent neuronal circuits emanating from restricted areas of the brain stem and projecting throughout the telencephalon. This common anatomical feature probably facilitates a coordinated, quasi-hormonal control of energy supply in large areas of brain.

The strong influence of small changes in the structure of agonists on their activity (Fig. 1) and the apparently competitive nature of the antagonism exerted by substances such as N,N -dimethyltryptamine (Fig. 2) suggest that serotonin-induced glycogenolysis is a receptor-mediated response. The pharmacol-

Table 1 Comparison of drug potencies on the hydrolysis of ^3H -glycogen in cortical slices and other models

Drugs	^3H -glycogen hydrolysis	K_{act} or K_i values (μM)		Serotonin-sensitive adenylate cyclase
		^3H -serotonin binding sites	^3H -spiperone binding sites	
Agonists				
Serotonin	20	0.004 (14)	2.7 (14)	1 (9) 0.001 (10)
Tryptamine	26	> 2 (12)		
5-methoxytryptamine	22	0.011 (14)	2.7 (14)	
N -methyltryptamine	9			
Bufofene	5	0.037 (14)		
(+)LSD	60	0.010 (14)	0.013 (14)	
Antagonists				
<i>a</i> , Serotonergic				
Metergoline (7 μM)	6	0.009 (17)	0.002 (17)	1 (9)
Methysergide (7 μM)	NE	0.088 (14)	0.003 (14)	0.2 (10)
Cyproheptadine (30 μM)	NE	1.5 (14)	0.002 (14)	0.2 (9) 0.6 (10)
N,N -dimethyltryptamine (2 μM , 5 μM)	0.8	0.17 (12)		
Ketanserin (30 μM)	NE	NE (15)	0.002 (15)	
Methiothepin (20 μM)	20	0.3 (17)	0.004 (17)	0.3 (9)
<i>b</i> , Neuroleptics				
Spiperone* (50 μM)	20	0.73 (14)	0.001 (14)	0.3 (11)
Pimozide* (3 μM)	5	> 4 (12)		
Pipamperone*	20	5 (15)	0.005 (15)	
Chlorpromazine* (3 μM)	4	6 (12)	0.02 (16)	0.4 (9)
<i>c</i> , Antidepressants				
Imipramine* (1 μM , 5 μM)	1	20 (17)	0.2 (17)	
Amitriptyline (0.5 μM)	1	5 (17)	0.01 (17)	
Mianserin (1 μM , 5 μM)	1.7	0.87 (14)	0.005 (14)	2.5 (11)
Doxepin (3 μM)	2.1			
Desimipramine (3 μM)	2.4	9 (17)	0.1 (17)	
Maprotiline* (3 μM)	1.6			
Inactive agents				
Alprenolol (100 μM), timolol (10 μM), mepyramine (30 μM), morphine (30 μM) and 5-hydroxyindoleacetic acid (100 μM), as either agonists or antagonists				

K_{act} values of compounds for ^3H -glycogen hydrolysis were obtained from the data of Fig. 1. K_i values were determined by one of two methods: (1) see Fig. 2. The rightward shift of the concentration-response curve to serotonin was determined in the presence of fixed concentration(s) of antagonists (concentrations indicated in parentheses) and the Cheng-Prusoff equation³⁵ was then applied. (2) See Fig. 2 legend. The effect of 20 μM serotonin was antagonized by agents (indicated by *) in increasing concentrations and their K_i values were calculated from their IC_{50} values using the equation: $K_i = \text{IC}_{50} (1 + S/K_a)$. As shown for N,N -dimethyltryptamine in Fig. 2 legend, the two methods give K_i values which are always in good agreement, and the average is given. Concentration-response curves were established using six to eight concentrations of serotonin, agonists or antagonists (for each concentration 8–20 distinct determinations were performed in 2–5 separate experiments). Numbers in parentheses refer to references from which data were taken. NE, non-effective agents at the indicated concentration.

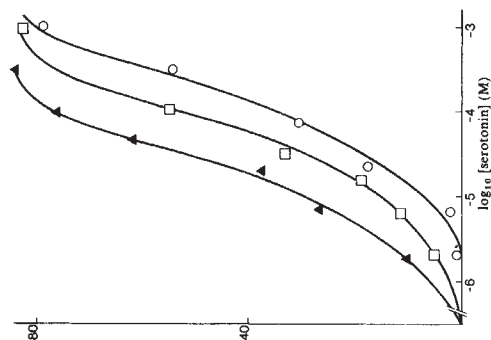


Fig. 2 Inhibition of the serotonin-induced glycogenolysis in slices from cerebral cortex by *N,N*-dimethyltryptamine. Following a 30 min preincubation in the presence of ^3H -glucose, serotonin was added at increasing concentrations and the incubation continued for a further 20 min. *N,N*-dimethyltryptamine was added 10 min before serotonin. Basal ^3H -glycogen level was $23.4 \pm 1.2 \text{ d.p.m.} \times 10^3$ per mg protein. Values are the mean of 10–15 determinations performed in 2–5 independent experiments. The K_{act} values of serotonin alone or in the presence of *N,N*-dimethyltryptamine were determined by computer³⁶ and a K_i value of $0.8 \mu\text{M}$ was obtained from the equation: $K_i = I / [(K_a/K_a') - 1]$, where K_a and K_a' are the K_{act} values of serotonin in the absence and presence of antagonist respectively, and I is the concentration of antagonist. In another experiment (not shown), using a fixed concentration of serotonin ($100 \mu\text{M}$) and increasing concentrations of *N,N*-dimethyltryptamine, the IC_{50} value of the latter compound was $10 \mu\text{M}$, from which a K_i value of $1.7 \mu\text{M}$ was calculated from the equation³⁵: $K_i = \text{IC}_{50} / (1 + S/K_a)$ where IC_{50} is the concentration of antagonist required to produce a 50% inhibition of the glycogenolytic effect of serotonin, S represents the concentration of serotonin and K_a is the K_{act} value of serotonin. ●, Serotonin; □, serotonin + $2 \times 10^{-6} \text{ M}$ *N,N*-dimethyltryptamine; ○, serotonin + $5 \times 10^{-6} \text{ M}$ *N,N*-dimethyltryptamine.

ological specificity of the receptors mediating the serotonin-induced glycogenolysis does not correspond, however, to that of other serotonergic receptors or putative receptors: for example, there are large differences in the apparent dissociation constants of the typical serotonergic antagonists regarding inhibition of glycogenolysis, adenylate cyclase or binding of ^3H -spiperone or ^3H -serotonin respectively (Table 1). This suggests that a novel subclass of serotonin receptors mediates the glycogenolytic response.

It is unlikely that these receptors represent the common site of action of the potent hallucinogenic indole derivatives which display relatively low apparent affinity for the receptors and behave as either agonists (tryptamine) or antagonists (*N,N*-dimethyltryptamine). In view of the low apparent affinity of neuroleptics (compared with their affinity for dopamine receptors), it is also unlikely that significant occupation of serotonin receptors mediating glycogenolysis occurs during antipsychotic treatments which result in body fluid levels of these drugs in the nanomolar range³¹. On the other hand, micromolar concentrations of tricyclic antidepressants belonging to various chemical classes all antagonize competitively the serotonin-induced response. The antagonist effect of these compounds seems unrelated to their ability to block the serotonin uptake process as, for example, compounds such as iprindole or desimipramine are almost inactive on this process^{32,33}. Because the effective concentrations of tricyclic compounds in the body fluids of patients undergoing antidepressant treatment seem to be in the micromolar range³⁴, it is likely that serotonin receptors mediating the glycogenolysis in brain are blocked during such treatments. It remains to be established whether this blockade participates in the antidepressant activity (or side effects) of these drugs.

Received 23 November 1981; accepted 21 May 1982.

1. Barchas, J. & Udsin, E. (eds) in *Serotonin and Behavior*, 1–316 (Academic, New York, 1973).
2. Fuxe, K., Schubert, J., Hokfelt T. & Jonsson, G. *Adv. biochem. Psychopharmac.* **10**, 67–74 (1974).
3. Green, A. R. & Grahame-Smith, *Nature* **260**, 487–491 (1976).
4. Aghajanian, G. K. & Wand, R. Y. in *Psychopharmacology: A Generation of Progress* (eds Lipton, M. A., Di Mascio, A. & Killam, K. F.) 171–193 (Raven, New York, 1978).
5. Besson, J. M. *Pain and Society* **17**, 161–182 (1980).
6. Garver, D. L. & Davis, J. M. *Life Sci.* **24**, 383–394 (1979).
7. Von Hungen, K., Roberts, S. & Hill, D. F. *Brain Res.* **84**, 257–267 (1975).
8. Enjalbert, A., Bourgoin, S., Hamon, M., Adrien, J. & Bockaert, J. *Molec. Pharmac.* **14**, 2–10 (1978).
9. Enjalbert, A., Hamon, M., Bourgoin, S. & Bockaert, J. *Molec. Pharmac.* **14**, 11–23 (1978).
10. Fillion, G. *et al. Life Sci.* **24**, 1813–1819 (1979).
11. Nelson, D., Herbert, A., Enjalbert, A., Bockaert, J. & Hamon, M. *Biochem. Pharmac.* **29**, 2445–2453 (1980).
12. Fillion G. M. B. *et al. Molec. Pharmac.* **14**, 50–59 (1978).
13. Nelson, D. L., Herbert, A., Bourgoin, S., Glowinski, J. & Hamon, M. *Molec. Pharmac.* **14**, 983–995 (1978).
14. Peroutka, S. J. & Snyder, S. H. *Molec. Pharmac.* **16**, 687–699 (1979).
15. Leysen, J. E. *et al. Life Sci.* **28**, 1015–1022 (1981).
16. Leysen, J. E. *J. Physiol., Paris* **77**, 351–362 (1981).
17. Peroutka, S. J., Lebovitz, R. M. & Snyder, S. H. *Science* **212**, 827–829 (1981).
18. Dolphin, A. C. & Greengard, P. *Nature* **289**, 76–79 (1981).
19. Quach, T. T., Rose, C. & Schwartz, J. C. *J. Neurochem.* **30**, 1335–1341 (1978).
20. Edwards, C., Nahorski, S. R. & Rogers, K. J. *J. Neurochem.* **22**, 564–572 (1974).
21. Schwartz, J. C., Martres, M. P., Rose, C., Quach, T. T. & Sokoloff, P. in *Chemical Neurotransmission, 75 Years* (eds Stjärne, L., Hedkvist, P., Lagercrantz, H. & Wonnmalm, A.) 411–423 (Academic, New York, 1981).
22. Quach, T. T., Duchemin, A. M., Rose, C. & Schwartz, J. C. *Molec. Pharmac.* **17**, 301–308 (1980).
23. Magistretti, P., Morrison, J. H., Shoemaker, W. J., Sapin, V. & Bloom, F. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6535–6539 (1981).
24. Sutherland, E. W. & Robison, G. A. *Pharmac. Rev.* **18**, 145–164 (1966).
25. Haylett, D. G. & Jenkinson, D. H. *J. Physiol., Lond.* **225**, 751–772 (1972).
26. Assimacopoulos Jeannet, F. O., Blackmore, P. F. & Exton, J. J. *biol. Chem.* **252**, 2662–2669 (1977).
27. Koizumi, J. *Prog. Histochem. Cytochem.* **6**, 1–37 (1974).
28. Dahlstrom, A. & Fuxe, K. *Acta physiol. scand.* **62**, 1–35 (1964).
29. Dahlstrom, A., Haggendal, J. & Atack, C. in *Serotonin and Behavior* (eds Barchas, J. & Udsin, E.) 87–96 (Academic, New York, 1973).
30. Garbarg, M., Barbin, G., Feger, T. & Schwartz, J. C. *Science* **186**, 833–835 (1974).
31. Seeman, P. *Pharmac. Rev.* **32**, 229–313 (1980).
32. Ross, S. B., Renyi, A. L. & Ogren, S. O. *Life Sci.* **10**, 1267–1277 (1971).
33. Maas, J. W. *Archs gen. Psychiat.* **32**, 1357–1361 (1975).
34. Glassman, A. H. & Perel, J. M. in *Psychopharmacology: A Generation of Progress* (eds Lipton, M. A., Di Mascio, A. & Killam, K. F.) 917–922 (Raven, New York, 1979).
35. Cheng, C. & Prussoff, W. H. *Biochem. Pharmac.* **22**, 3099–3108 (1973).
36. Parker, R. B. & Waud, D. R. *J. Pharmac. exp. Ther.* **177**, 1–24 (1971).

Monoclonal antibodies against human T lymphocytes label Purkinje neurones of many species

J. A. Garson, P. C. L. Beverley,* H. B. Coakham & E. I. Harper

Brain Tumour Research Laboratory, Departments of Neuropathology and Neurosurgery, Frenchay Hospital, Bristol BS16 1LE, UK

*ICRF Human Tumour Immunology Group, Faculty of Clinical Sciences, School of Medicine, University Street, London WC1E 6JJ, UK

Conventional antibodies have long been used in an attempt to produce specific neural markers^{1,2}. Such markers would be invaluable for studying the structural organization and development of the nervous system. Unfortunately, they have not generally been found to discriminate satisfactorily between neuronal subpopulations^{3,4}. Recent developments of the hybridoma technique⁵, however, promise to provide monoclonal antibodies of adequate specificity. Such antibodies have already been generated and shown to be capable of distinguishing between individual neurones of the leech nervous system⁶. We report here two monoclonal antibodies which, although generated against human T cells^{7,8}, react exclusively with Purkinje cells in the vertebrate central nervous system (CNS). This new specificity arose out of a fortuitous observation made during examination of the lymphocyte infiltration of human cerebellar tumours with the monoclonal antibody, UCHT1. Although widely used as a T-cell marker, its reaction with neural tissue has not hitherto been described. To our knowledge, this is the first description of a monoclonal antibody which recognises a functionally discrete neuronal population in the vertebrate brain.

UCHT1 is an IgG1 mouse monoclonal antibody, derived from a fusion between P3/NS1/1-Ag4-1 myeloma cells and splenic lymphocytes from BALB/c mice immunized with human infant thymocytes and Sezary cells. The antibody had been previously characterized in detail⁷ and was thought to bind exclusively to an antigen of 22,000 molecular weight (MW) present on the surface of human T lymphocytes. Therefore, we were surprised that it produced strong staining of Purkinje cells, by indirect immunofluorescence, when applied to frozen sections of human cerebellum.

Purkinje cells provide the only efferent pathway from the cerebellar cortex, their axons terminating chiefly in the deep cerebellar nuclei⁹. Their large flask-shaped cell bodies and extensive dendritic arborizations give them a unique appearance. Figure 1 shows intense cytoplasmic staining of Purkinje cells by UCHT1, extending throughout axons and dendrites. Note that cerebellar granule, stellate, basket and glial cells fail to stain. Human, mouse, rat, guinea pig and quail brains produced identical Purkinje cell staining. No other cells were stained in the cerebrum, cerebellum or spinal cord of any of the species tested. Extensive screening of other non-lymphoid tissues revealed no further cross-reactivity. The specificity of UCHT1 staining is summarized in Table 1. Absorption experiments (Table 2) confirmed restriction of the UCHT1 neural antigen to the cerebellum. The discovery of Purkinje cell labelling by UCHT1 led us to investigate other anti-haemopoietic cell monoclonal antibodies for similar cross-reactivity (see Table 3). Of 12 reagents tested, only one (Leu-4; Becton Dickson) had anti-Purkinje cell activity. It produced cytoplasmic staining which was indistinguishable from that of UCHT1. Cross blocking experiments on human T lymphocytes (unpublished data) indicate that UCHT1, Leu-4 and another pan T-cell marker, OKT3¹⁰ (Ortho Pharmaceuticals), recognize closely related antigenic determinants. It is, therefore, of special interest to note that OKT3 does not react with Purkinje cells.

Table 1 UCHT1 reactivity

Positive	Human	Negative	Other species
Human T cells	Erythrocytes	Rat:	
Human Purkinje cells	Polymorphs	Cerebellum†	
Mouse Purkinje cells	Monocytes	Olfactory bulb	
Guinea pig Purkinje cells	Platelets	Cerebrum	
Rat Purkinje cells*	Non T lymphocytes	Midbrain	
Quail Purkinje cells	Cerebrum	Brain stem	
	Cerebellum†	Spinal cord	
		Dorsal root ganglion	
	Spinal cord	Spleen	
	Pituitary	Kidney	
	Muscle	Lung	
	Liver	Bladder	
	Kidney	Tongue	
	Skin	Oesophagus	
	Breast	Duodenum	
	Thyroid	Aorta	
	Adrenal	Mouse:	
		Cerebellum†	
	Colon	Cerebrum	
	Oesophagus	Spinal cord	
	Stomach	Kidney	
	Tongue	Guinea pig:	
		Cerebellum†	
	Buccal mucosa	Spleen	
	Connective tissue	Quail:	
		Cerebellum†	
	Primary cerebral tumours (25)	Cerebrum	
	Secondary cerebral tumours (8)	Spinal cord	

* Adult and 10-day-old neonates were tested.

† Cerebellum negative apart from Purkinje cells.

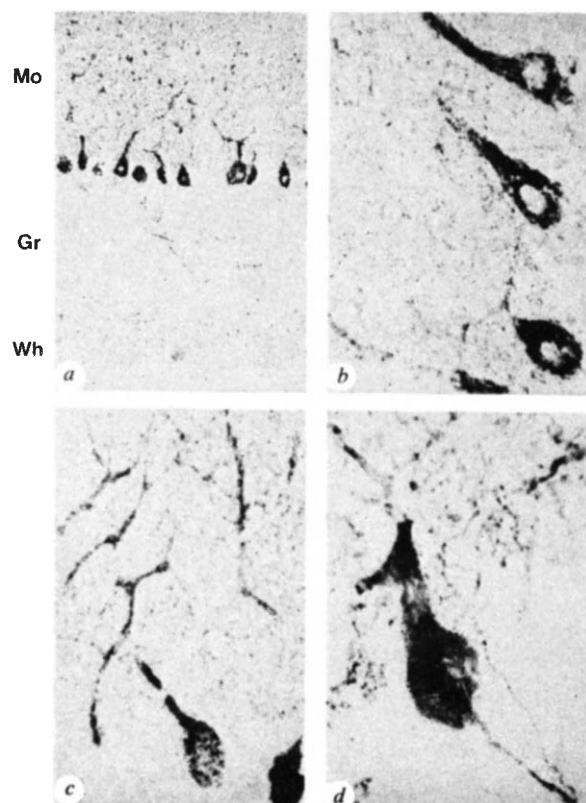


Fig. 1 Guinea pig Purkinje cells stained by UCHT1 using the indirect immunoperoxidase technique. Frozen sections (6 μ m) of cerebellum were fixed briefly in acetone then incubated at 20 °C for 30 min with a purified IgG fraction of UCHT1 ascites, diluted 1:200. The slides were washed for 10 min in phosphate-buffered saline then incubated at 20 °C for 30 min with peroxidase-conjugated rabbit anti-mouse immunoglobulin (Dako), diluted 1:100. Diaminobenzidine (0.6 mg ml⁻¹) was used as the chromogenic substrate. *a* Shows Purkinje cell axons traversing the granule cell layer (Gr) to join other Purkinje axons of the cerebellar white matter (Wh). These axons were seen to terminate on neurones of the deep cerebellar nuclei. Such neurones showed no staining with UCHT1. Purkinje cell branching dendrites in the molecular layer (Mo) are clearly stained. *b*, The antigen is seen to be cytoplasmic, as the nuclei show no staining. *c* Shows the dendritic arborizations of the molecular layer; *d* shows an axon as it leaves the base of a Purkinje cell body.

The Purkinje cell antigen does not survive fixation in formaldehyde, paraformaldehyde, glutaraldehyde, osmium tetroxide, Carnoy's or Bouin's fixatives but is unaffected by fixation in acetone at room temperature for 5 min. The antigen is best demonstrated in fresh material but can be detected in sections stored for up to 3 days at -20 °C. Beyond this time, staining becomes unpredictable.

The relatively labile nature of the Purkinje antigen has so far prevented any determination of its molecular weight or biochemical characteristics. Without such data it is unsafe to assume that the UCHT1-defined Purkinje cell and T-cell antigens are identical^{11,12}. Indeed, their differing species distributions and different locations within the cell (cytoplasm versus plasma membrane) suggest that they may be unrelated molecules which merely share a common antigenic determinant.

Although the nature of UCHT1-defined Purkinje specific antigen is unknown, it does not seem to match any of the other previously reported Purkinje cell-associated proteins such as P-400 (ref. 13). This 400,000 MW protein is not detectable in the Purkinje cell axon and is therefore unlikely to be the antigen recognized by UCHT1. Prostaglandin dehydrogenase¹⁴ and the calcium binding protein parvalbumin¹⁵ have also been reported to be Purkinje cell-related but lack the highly restricted localization of the UCHT1-defined antigen. The intestinal peptide hormone motilin has recently been detected in Purkinje cells¹⁶

Table 2 Absorption experiments with UCHT1

	T lymphocyte staining	Purkinje cell staining
UCHT1 unabsorbed	++	++
UCHT1 cerebrum absorbed	++	++
UCHT1 cerebellum absorbed	±*	—

UCHT1 was diluted to 1:1200 (one doubling dilution below end titre) and absorbed with homogenates of rat cerebrum and cerebellum. Absorption was carried out at 20°C for 1 h with equal volumes of diluted antibody and homogenate. Following absorption, UCHT1 binding on human T cells and Purkinje cells in frozen section was visualized by indirect immunofluorescence using a fluorescein-conjugated goat anti-mouse immunoglobulin, diluted 1:80.

* Weak residual staining of T cells.

but, unlike the UCHT1-defined antigen, is only present in a proportion of Purkinje cells. Furthermore, UCHT1 fails to stain motilin-containing cells of the rat duodenum. Finally, cyclic GMP-dependent protein kinase has been shown, within the rodent nervous system, to be restricted almost exclusively to Purkinje cells¹⁷. However, in contrast to the UCHT1 antigen, the enzyme is also present in high concentration in the muscular wall of the aorta and in various other smooth muscle cells (see Table 1).

Purkinje-specific antigens have previously been demonstrated by immunological studies with two conventional polyclonal heteroantisera. The first¹⁸, although operationally specific, was only effective at very low titre. The second¹⁹, raised against a glycoprotein extracted from rat cerebellum, appeared Purkinje cell-specific, but specificity was demonstrated by absorption rather than by direct testing so that small populations of antigen-containing cells elsewhere in the nervous system might have been overlooked. In addition to these studies with polyclonal antisera, a monoclonal antibody, designated CE5, reactive with Purkinje cells, has also been described²⁰. CE5, unlike UCHT1, however, is relatively nonspecific, reacting with many different types of neurone including dorsal root ganglion cells and pyramidal neurones; it also labels cardiac muscle and the protozoan *Trypanosoma cruzi*. The high titre, high specificity and unlimited supply of UCHT1 and Leu-4 make them ideal immunological markers for Purkinje cells.

The work described here, together with preliminary experiments on neonatal mouse cerebellum and cerebellar mutant strains, suggests that UCHT1 and Leu-4 will become invaluable tools for studying the organization and development of cerebellar architecture and connectivity.

This work was financed by the MRC, the Imperial Cancer Research Fund, Bristol Brain Tumour Research Fund and the

Francis and Augustus Newman Foundation. We thank Dr D. B. Brownell, D. Rowe, T. Gradidge and J. Singh for their help and K. Caddy for providing the data on fixatives. Monoclonal antibody 308 was donated by J. Kemshead, RFB-1 by G. Janossy and CR3/43 by D. Y. Mason. OKT3, OKT4 and OKT8 were supplied by Ortho Pharmaceuticals and Leu-4 and Leu-2a donated by Becton Dickinson.

Received 22 March; accepted 18 May 1982.

1. Raff, M. C. *et al.* *Brain Res.* **174**, 283–308 (1979).
2. Schachner, M., Wortham, K. A. & Kincade, P. W. *Cell Immun.* **22**, 369–374 (1976).
3. Mirsky, R., Wendon, L. M. B., Black, P., Stolkin, C. & Bray, D. *Brain Res.* **148**, 251–259 (1978).
4. Rutishauser, U., Gall, W. E. & Edelman, G. M. *J. Cell Biol.* **79**, 382–393 (1978).
5. Kohler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
6. Zipser, B. & McKay, R. *Nature* **289**, 549–554 (1981).
7. Beverley, P. C. L. & Callard, R. E. *Eur. J. Immun.* **11**, 329–334 (1981).
8. Ledbetter, J. A., Frankel, A. E., Herzenberg, L. A. & Herzenberg, L. A. in *Monoclonal Antibodies and T Cell Hybridomas* (eds Hammerling, U., Kearney, J. & Hammerling, G.) (Elsevier, New York, in the press).
9. Eccles, J. C. in *Cerebro-Cerebellar Interaction* (eds Massion, J. & Sasaki, K.) (Elsevier, New York, 1979).
10. King, P. C., Goldstein, G., Reinhertz, E. L. & Schlossman, S. F. *Science* **206**, 347–349 (1979).
11. Dulbecco, R. *et al.* *Nature* **292**, 772–774 (1981).
12. Lane, D. & Koprowski, H. *Nature* **296**, 200–202 (1982).
13. Mikoshiba, K., Huchet, M. & Changeux, J. P. *Dev. Neurosci.* **2**, 254–275 (1979).
14. Siggins, G., Hoffer, B. & Bloom, F. *Ann. N.Y. Acad. Sci.* **180**, 302–319 (1971).
15. Celio, M. R. & Heizmann, C. W. *Nature* **293**, 300–302 (1981).
16. Nilaver, G., Defendini, R., Zimmerman, E. A., Beinfeld, M. C. & O'Donohue, T. L. *Nature* **295**, 597–598 (1982).
17. Walter, U., DeCamilli, P., Lohmann, S. M., Miller, P. & Greengard, P. *Cold Spring Harb. Conf. Cell Proliferation* **8**, 141–157 (1981).
18. Woodhams, P., Mallet, J., Changeux, J. P. & Balazs, R. *Dev. Biol.* **72**, 320–326 (1979).
19. Reeber, A., Vincendon, G. & Zanetta, J. P. *Brain Res.* **229**, 53–65 (1981).
20. Wood, J. N., Hudson, L., Jessell, T. M. & Yamamoto, M. *Nature* **296**, 34–38 (1982).
21. Oger, J., Szuchet, S., Antel, J. & Arnason, B. *Nature* **295**, 66–68 (1982).

Germinal centre B cells: antigen specificity and changes in heavy chain class expression

G. Kraal, I. L. Weissman & E. C. Butcher

Laboratory of Experimental Oncology, Department of Pathology, Stanford University Medical Centre, Stanford, California 94305, USA

Germinal centres are histologically defined aggregates of blast cells that occur in B-cell areas of lymphoid tissues after antigenic stimulation. They are believed to be associated with the development of B-cell memory and plasma cell (especially secondary, IgG and IgA) responses^{1–3}. Recent studies of murine lymphoid tissues have defined cell-surface markers that distinguish germinal centre B cells from other mature B cells^{4–7}, permitting their identification and characterization in cell suspensions. Here we have used these markers to define and study germinal centre cells in lymph nodes, and have found that they constitute a unique population of B cells which (1) arises in response to antigenic stimulation, (2) contains nearly all of the demonstrably antigen-specific B cells in the stimulated organ, (3) bears surface IgM after primary stimulation and (4) as a population, demonstrates isotype switching to a predominant surface IgG phenotype after secondary stimulation with specific antigen. These findings demonstrate that germinal centres are a major site of proliferation and differentiation of antigen-specific B cells *in vivo*, and suggest that the germinal centre microenvironment may have an important role in heavy chain class switching during B-cell responses.

Germinal centre B cells possess unique cell-surface features. One of the most important of these is defined by peanut agglutinin (PNA)^{4–7}, a plant lectin with specificity for terminal galactosyl residues. In antigen-stimulated lymphoid organs, fluorescence-labelled PNA defines a discrete population of lymphocytes (termed PNA^{hi}) that bind 10–30 times as much PNA as the major population of lymphocytes⁷ (PNA^{lo}). Immunohistological^{4,5,7}, electron microscopic⁵ and cell suspension^{4–7} studies of mouse Peyer's patches and stimulated lymph nodes have shown that the B cells in the PNA^{hi} population are almost exclusively germinal centre lymphocytes. B cells

Table 3 Reactivity of anti-haematopoietic cell monoclonal antibodies with Purkinje cells

Monoclonal antibody	Haematopoietic cell reactivity	Purkinje cell staining
UCHT1	Pan T cell	+
Leu-4	Pan T cell	+
OKT3	Pan T cell	—
OKT4	Helper T cells	—
OKT8*	Suppressor/cytotoxic T cells	—
UCHT2	Pan T cell + B.CLL	—
UCHT3	Pan T cell + B.CLL	—
CR3/43	Anti HLA-Dr	—
Leu-2a	Suppressor/cytotoxic T cells	—
2D1	All leukocytes	—
TG1	Granulocytes	—
308	Granulocytes	—
RFB-1	Lymphoid stem cells	—

* OKT8 has recently been shown to exhibit anti-ovine oligodendrocyte cross-reactivity *in vitro*²¹. However, we have been unable to demonstrate OKT8 immunostaining of oligodendrocytes in frozen sections of human white matter, optic nerve, multiple sclerosis plaques or oligodendroglioma.

outside germinal centres, as well as plasma cells and most T cells, bind only low levels of PNA^{5,7}. A small proportion of T cells may also be PNA^{hi} (refs 7, 8). A secondary independent marker is IgD: immunohistological^{9,10} and flow cytometric⁷ studies have shown that germinal centre lymphoid cells lack surface IgD, whereas essentially all other mature B cells are IgD⁺ (ref. 7). Thus, germinal centre B cells can be identified confidently in cell suspensions of lymph nodes and Peyer's patches by virtue of their unique PNA^{hi}, IgD⁻, Thy-1⁻ phenotype.

BALB/c or BALB.B mice were immunized in the front foot pads with heterologous erythrocytes, and cell suspensions were made from brachial and axillary lymph nodes during or near the peak of germinal centre responses (6 days after primary or secondary immunization). Several studies were carried out to confirm the usefulness of PNA in identifying germinal centre lymphoid cells in these lymph nodes. Fluorescence activated cell sorter (FACS) analyses demonstrated the appearance of a discrete population of PNA^{hi} cells in the stimulated nodes, constituting up to 10–12% of the lymph node cells (and up to 20–30% of B cells) 6 days after primary immunization (see Fig. 1). Double fluorescence microscopy of frozen sections of stimulated nodes stained with tetramethylrhodamine isothiocyanate (TRITC)-conjugated PNA and counterstained with fluorescein isothiocyanate (FITC)-rabbit anti-mouse Ig confirmed the remarkable specificity, within the B-cell lineage in this organ, of intense PNA staining for germinal centre cells (not shown, see Raedler *et al.*⁵ for similar findings). Staining of FACS-separated PNA^{hi} and PNA^{lo} populations confirmed the observa-

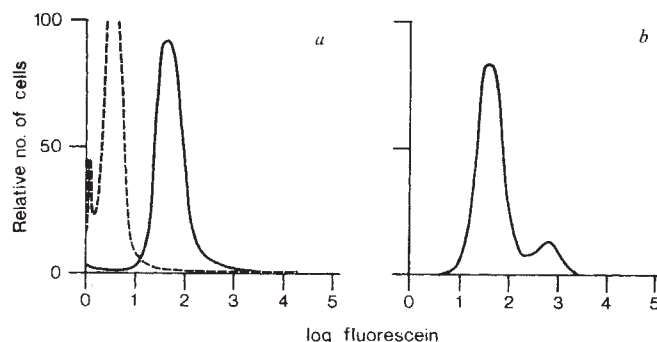


Fig. 1 Axillary and brachial lymph node cells from BALB/c mice were incubated for 5 min in an excess of FITC-conjugated PNA (EY Laboratories) at $100 \mu\text{g ml}^{-1}$ in phosphate-buffered saline, washed, and analysed on the FACS (Becton Dickinson). *a*, Unstimulated lymph node cells were stained with FITC-PNA in the presence (dotted line) or absence (solid line) of 0.2 M galactose. Most lymphocytes bind low levels of PNA to galactose-inhibitable sites, and are therefore referred to as PNA^{lo}. *b*, Six days after footpad immunizations with $1-2 \times 10^7$ SRBC, a discrete population of PNA^{hi} cells has arisen which binds about 20 times as much PNA as the PNA^{lo} lymphocytes. To determine if this PNA^{hi} population included significant numbers of plasma cells or plasmablasts, cytocentrifuge preparations of FACS-sorted PNA^{hi} and PNA^{lo} populations were air-dried, fixed 10 min in cold acetone, counterstained with rabbit anti-mouse immunoglobulin and examined by fluorescence microscopy. Cells containing cytoplasmic immunoglobulin segregated predominantly with the PNA^{lo} population. Of 400 sorted PNA^{lo} cells counted, 41 contained detectable internal immunoglobulin; 27 of these were mature plasma cells with eccentric nuclei, and the remaining 14 had the appearance of plasmablasts. Of 1,000 PNA^{hi} sorted cells examined, only one apparent plasma cell and seven plasmablasts were identified.

Table 1 The phenotype of PNA^{hi} sample cells in lymph nodes after immunization

Antibody directed against	% Of cells stained	
	Primary response	Secondary response
B cells (B220)	94	83
Total immunoglobulin	84	80
IgM	62	25
IgG	9	60
IgG + γ blockade	ND	1
IgG + α and μ blockade	ND	64
IgA	4	3
IgD	1	1
T cells (Thy-1 + Lyl-1)	9	10
Rabbit anti-KLH	0	1
RA3-2C1	0	0

BALB.B mice, 6–8 weeks of age, were injected with $1-2 \times 10^7$ SRBC in the front footpads. For analysis of the primary response, axillary and brachial lymph nodes were removed 6 days after immunization. For the secondary response lymph nodes were collected 6 days after a second injection of antigen with a 2-week interval between first and second injection. Cells were stained at a density of 10^6 cells per $10 \mu\text{l}$ of antiserum at saturating concentrations for 5–10 min and washed between first and second stages by centrifugation through serum. After incubating the cells in the second stage for 5 min, an equal volume of TRITC-conjugated PNA (EY Laboratories) at $100 \mu\text{g ml}^{-1}$ was added for another 3–4 min. The cells were then directly suspended in 1–2 ml of 2% formalin in phosphate-buffered saline washed through serum and examined microscopically. Affinity-purified anti-mouse total immunoglobulin, anti-keyhole limpet haemocyanin (KLH), and μ -chain-specific¹⁶ and α -chain specific⁷ antibodies (gifts from Drs C. Nottenburg and R. L. Coffman) and an Fab preparation of an anti- γ -serum¹⁷ (a gift of Dr E. Vitetta) were produced in rabbits and used in combination with a fluoresceinated sheep anti-rabbit immunoglobulin second stage (Institut Pasteur, Paris). To detect δ chains a biotinylated monoclonal anti-IgH-5a, H10.4.22¹⁸, was used in combination with fluoresceinated avidin (EY Laboratories). B cells were detected by a newly developed monoclonal rat antibody (RA3-6B2, a gift of Dr R. L. Coffman) specific for B220, the B-lineage-specific form of the T200 family of molecules¹³. A mixture of monoclonal anti-Lyl-1 (53-7.3) and anti-Thy-1 (30-H12)¹⁹ was used to detect T cells with a highly specific fluoresceinated rabbit anti-rat immunoglobulin second stage. A monoclonal rat antibody (RA3-2C1) with no known specificity was used as a control. The specificity of the rabbit anti-mouse γ chain reagent was tested by attempting to block staining either with a pool of IgG myeloma proteins or with a mixture of IgA and IgM myeloma proteins. Preincubation of the antiserum for 15 min with mouse IgGs totally inhibited staining but an excess of IgA (TEPC 15) and IgM myeloma proteins had no effect. 150–200 PNA^{hi} cells (with the characteristic large size and/or irregular outline of germinal centre cells) were scored, single-blind, using a Zeiss fluorescence microscope with exciter/barrier filter combinations for rhodamine and fluorescein. The data are pooled from two experiments with similar results and are expressed as the % of sample cells staining.

tion, previously based on immunohistological studies^{5,7}, that plasma cells and plasmablasts are PNA^{lo} (see Fig. 1 legend).

For studies of the surface phenotype of the PNA^{hi} population, cells were labelled with TRITC-conjugated PNA and counterstained with various (fluorescein-detected) monoclonal antibodies or heterologous antisera. Putative germinal centre cells were identified microscopically by their intense staining with PNA in combination with their characteristic larger size and frequently irregular cell outline. These criteria excluded most PNA^{hi} T cells, yet included nearly all B cells with the characteristic IgD⁻ phenotype of germinal centre cells. Cells so identified were then examined under fluorescein filters for staining with immunological reagents.

Initial experiments describing the surface phenotype of germinal centre cells after primary and secondary immunizations are presented in Table 1. Only 10% of the sample PNA^{hi} population were contaminating T cells; the remainder were IgD⁻ B cells. The results demonstrate that IgM is the predominant isotype of surface immunoglobulin during the primary response, but in a secondary response, germinal centre cells mainly express surface IgG. Although the present findings do not demonstrate that these B cells synthesize their surface immunoglobulin, we have previously presented evidence based on allelic exclusion that the comparable population of Peyer's patch germinal centre cells do synthesize their major class of surface immunoglobulin (IgA)^{6,7}. (In striking contrast to the pattern of immunoglobulin expression by peripheral node germinal centre cells, 50–75% of germinal centre cells in Peyer's patches express surface IgA⁷.) These findings suggest that germinal centres may have an important role in the process of isotype switching.

We wished therefore to determine if the B cells proliferating in germinal centres are specific for the immunizing antigen, as suggested by work of Sordat *et al.*²¹, or if germinal centres merely represent sites of nonspecific expansion and perhaps isotype switching of all available B cells. Kanowitz-Klein *et al.*¹² have shown that the proportion of antigen binding B cells bearing surface IgG increases during the immune response to heterologous erythrocytes. Thus the switch we observed from a predominance of IgM⁺ to IgG⁺ cells in germinal centres with restimulation with the same antigen suggested that these cells

Table 2 Heavy chain isotype switching of germinal centre cells is antigen specific

1° Injection	Antigen	2° Injection	% Of PNA ^{hi} cells with surface	
			IgM	IgG
SRBC		SRBC	31	48
SRBC		HuRBC	56	13
SRBC		None	46	14
HuRBC		SRBC	52	21
HuRBC		HuRBC	36	49

BALB.B mice were immunized in the front foot pads with either SRBC or HuRBC. After 3 weeks the animals were boosted with one of the two antigens. Five days later lymph node cells were prepared and stained as described in Table 1 legend. 150–200 PNA^{hi} cells were counted single-blind. The results are from a representative experiment, expressed as the percentage of PNA^{hi} sample cells staining with the antiserum.

might be antigen specific. This reasoning was greatly strengthened by the demonstration that this IgM to IgG switch was itself antigen specific. After priming with SRBC, secondary immunization with SRBC reproducibly caused a switch in the germinal centre cells to a predominance of IgG⁺ cells, whereas secondary injection of human red cells (HuRBC) caused no shift (see Table 2). After HuRBC priming, secondary HuRBC but not SRBC injection caused the switch.

More direct evidence of antigen specificity has been obtained at the single cell level. By using rosetting techniques, we have found that almost all specific-antigen-binding B cells have the surface phenotype of germinal centre cells during the active immune response to SRBC. Suspensions of immunized lymph nodes were incubated with RBC briefly on ice, and subsequently stained with TRITC-PNA. Ten per cent of PNA^{hi}, IgD⁺ cells in SRBC-immunized lymph nodes formed SRBC rosettes. As many SRBC antigens are probably either located on the erythrocyte surface, or not sufficiently represented to support rosette formation, this proportion of rosetting cells probably underestimates the percentage of germinal centre cells that are antigen specific. Furthermore, after removal of glass-adherent cells (presumably macrophages binding cytophilic antibody), 98% (104/106) or rosetting cells were PNA^{hi}, and 100% (100/100) stained with RA3-6B2, a monoclonal antibody against B220¹³, the B-lineage specific form of the T200 family of molecules. Similar results were obtained after secondary injections.

Rosette formation was antigen-specific. Six days after foot pad injection of SRBC or HuRBC, immunized lymph node

Table 3 Germinal centre cells form rosettes specifically with the immunizing antigen

Immunogen	No. of rosetting cells binding		
	HuRBC	HuRBC and SRBC	SRBC
SRBC	0	0	100
HuRBC	113	0	0
SRBC and HuRBC	51	0	75

BALB.B mice were injected with SRBC, HuRBC or a 1:1 mixture of both erythrocyte types. After 6 days cell suspensions from axillary and brachial lymph nodes were prepared and incubated for 1 h at 37°C in glass tubes to remove adherent cells. Nonadherent cells were then added to a 1:1 mixture of SRBC and HuRBC and allowed to form rosettes for 10 min at 4°C. The cell mixtures were spun down and resuspended with rhodaminated PNA for 5 min after which they were diluted and directly examined in the fluorescence microscope. SRBC were labelled by incubation with FITC before rosette formation²⁰. Rosettes were (1) first identified microscopically under dark-field illumination, (2) then examined under UV epi-illumination using the filter combination for rhodamine to determine if the rosetting cells were PNA^{hi}, and if so, (3) the type of erythrocyte in the rosette was determined using filters for fluorescein. Rosettes were scored when at least 10 erythrocytes surrounded one nucleated centre cell. Approximately 10–15% of the specific rosettes, thus defined, contained one or two heterologous contaminating erythrocytes, but the frequency was no greater in recipients of mixed erythrocytes than in animals immunized with only one type. All rosette-forming cells were B cells (100 RA3-6B2⁺ per 100 rosettes counted); nearly all were PNA-binding (104 PNA^{hi} per 106 rosettes counted); and 10% (9.3; 10.1%) of PNA^{hi} cells formed rosettes. Similar results were obtained after secondary immunization. Truly mixed rosettes could be easily identified using this technique. In recipients immunized with both RBC types many mixed rosettes were observed before removal of adherent cells (presumably macrophages bearing cytophilically absorbed antibody).

cells were incubated with a mixture of HuRBC, and FITC-labelled SRBC. As shown in Table 3, PNA^{hi} cells of SRBC recipients only rosetted with SRBC, whereas only HuRBC rosettes were formed after HuRBC immunization. To rule out the possibility that these results were simply due to cytophilic adsorption of anti-RBC antibodies, individual mice were immunized with a mixture of HuRBC and SRBC in the same footpad, and 6 days later the specificity of PNA^{hi} rosetting cells was determined. If cytophilic antibodies were involved, we could expect most rosettes to be mixed, containing both FITC-SRBC and unlabelled HuRBC. Instead, all rosettes were formed specifically with one or the other RBC type, providing strong evidence that PNA^{hi} cells form rosettes by synthesis and surface expression of specific antibody.

The present findings support the concept that antigen-specific virgin B cells are recruited in the formation of germinal centres during a primary immune response, and subsequently undergo differentiation and selection in this site. Although many differentiation events may occur within the germinal centre population (for example, affinity maturation, suppression and/or reprogramming of normal lymphocyte migration mechanisms⁶), we have focused initially on changes in heavy chain class expression. If the precursors of germinal centre cells are IgD⁺, IgM⁺ lymphocytes (as seems likely, although other lineages cannot be ruled out^{11,14}), the first such alteration in isotype expression, which may be necessary on entry into the germinal centre, is the loss of surface IgD. Occurrence and significance of IgD regulation however, remain speculative. More interesting is the direct demonstration of changes in the proportions of germinal centre cells expressing IgM and IgG during the immune response. The switch to a predominance of IgG bearing cells, and the observation that almost all antigen-specific B cells seem to be maintained in the PNA^{hi} pool, strongly argue that the genetic rearrangements leading to heavy chain class switching occur in the germinal centre. The specific lack of IgA expression in peripheral lymph node as opposed to Peyer's patch germinal centres suggests that the germinal centre microenvironment (perhaps in conjunction with T-cell influences¹⁵) may be capable of directing the heavy chain class switch, or perhaps of selecting cells that have randomly switched to the isotype appropriate to the site of the immune response. Although our findings support a role for germinal centres in the production of memory B cells and plasma cell precursors¹⁻³, a thorough understanding of the role of germinal centres in humoral immunity must await direct lineage analyses.

This work was supported by NIH grant AI 09072. G.K. is the recipient of a Research Fellowship award from the Fogarty International Center. E.C.B. is a Senior Fellow of the American Cancer Society.

Received 9 November 1981; accepted 5 May 1982.

- Thorbecke, G. J., Romano, T. J. & Lerman, S. P. in *Prog. Immun.* **II** 3, 25–34 (1974).
- Nieuwenhuis, P., Gastkemper, N. A. & Opstelten, D. *Ciba Fdn Symp.* **84**, 246–264 (1981).
- Klaus, G. G. B. & Kunkl, A. *Ciba Fdn Symp.* **84**, 264–280 (1981).
- Rose, M. L., Birbeck, M. S. C., Wallis, V. J., Forrester, J. A. & Davies, A. J. S. *Nature* **284**, 364–366 (1980).
- Raedler, A., Raedler, E., Arndt, R. & Thiele, H. G. *Immun. Lett.* **2**, 335–338 (1981).
- Butcher, E. C., Reichert, R. A., Coffman, R. L., Nottenburg, C. & Weissman, I. L. *Adv. exp. Med. Biol.* (in the press).
- Butcher, E. C. et al. *J. Immun.* (submitted).
- London, J., Berrih, S. & Bach, J. F. *J. Immun.* **121**, 438–443 (1978).
- Martin, L. N. & Leslie, G. A. *Immunology* **33**, 865–872 (1977).
- Stein, H. et al. *J. Histochem. Cytochem.* **28**, 746–760 (1980).
- Jacobson, E. B. et al. *J. exp. Med.* **154**, 318–322 (1981).
- Kanowitz-Klein, S., Vitetta, E. S., Korn, E. L. & Ashman, R. F. *J. Immun.* **122**, 2349–2355 (1979).
- Coffman, R. L. & Weissman, I. L. *Nature* **289**, 681–683 (1981).
- Opstelten, D., Stikker, R., Van der Heyden, D. & Nieuwenhuis, P. *Virchows Arch. path. Anat. Physiol.* **34**, 53–62 (1980).
- Elson, C. O., Heck, J. A. & Strober, W. *J. exp. Med.* **149**, 632–643 (1979).
- Coffman, R. L. & Cohn, M. *J. Immun.* **118**, 1806–1815 (1977).
- Zan-bar, I., Strober, S. & Vietta, E. S. *J. Immun.* **123**, 925–930 (1979).
- Oi, V. T. P., Jones, P., Goding, S. W., Herzenberg, L. A. & Herzenberg, L. A. *Curr. Topics Microbiol. Immun.* **81**, 115–129 (1978).
- Ledbetter, J. A. & Herzenberg, L. A. *Immun. Rev.* **47**, 63–90 (1979).
- Butcher, E. C. & Weissman, I. L. *J. Immun. Meth.* **37**, 97–108 (1980).
- Sordat, B., Sordat, M., Hess, M. W., Stoner, R. D. & Cottier, H. *J. exp. Med.* **131**, 77–92 (1970).

Somatic variants of murine immunoglobulin λ light chains

Alfred L. M. Bothwell*†, Michael Paskind*†, Michael Reth‡, Thereza Imanishi-Kari†, Klaus Rajewsky‡ & David Baltimore*†

* Whitehead Institute for Biomedical Research, and

† Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡ Institut für Genetik der Universität zu Köln, 5000 Köln 41, FRG

Studies of the murine λ 1 light chains produced by myeloma cells provided the first evidence for somatic point mutation of germ-line variable (V) region genes. An examination of the variable regions of 19 λ 1 chains revealed seven which differed from a common sequence by one to three amino acid substitutions¹. Subsequently, one of these presumed somatic variants of the single λ 1 V gene was characterized by DNA sequence analysis of the rearranged functional gene². The predicted DNA sequence alteration was observed and no silent mutation was evident. These studies of λ chain variants suggested that the hypervariable, complementarity-determining regions (CDRs) might be a preferred site of somatic mutation because all seven characterized variants contained substitutions only in these regions. By contrast, comparisons of closely related κ chain variable region amino acid sequences, and more recently V_{κ} and V_H genes⁸⁻¹², have suggested that somatic mutation probably occurs in codons for both framework and CDR residues. To examine this apparent discrepancy between the sites of somatic mutation in λ and κ genes, we have determined the nucleotide sequence of two λ 1 gene from hybridomas and a λ 2 gene from a myeloma. These sequences demonstrate that somatic mutation in λ genes can occur in both the framework and CDR residues.

We have previously characterized the sequences of genes for the V_H regions of two antibodies which bear (4-hydroxy-3-nitrophenyl)acetyl (NP)^b idiotypic determinants⁸. These antibodies were from hybridomas of C57BL/6 origin and possessed λ 1 light chains: one was from hybridoma B1-8 which secretes an IgM antibody whose μ chain is specified by a germ-line nucleic acid sequence; another was from the S43 hybridoma which produces an IgG2a antibody in which 10 base pairs (bp) in the V_H region were altered from the germ-line configuration by somatic point mutations. We determined, therefore, whether somatic mutation in the λ 1 light chains paralleled that observed in their associated V_H regions.

In addition, we have determined the nucleotide sequence of p λ 2-1, the λ 2 cDNA clone derived from MOPC-315¹³. A comparison of the amino acid sequence of this protein with the sequence predicted from the germ-line λ 2 V gene indicated the presence of somatic mutation¹⁴, which was further defined by determining the nucleotide sequence encoding this protein.

Molecular cDNA clones of the λ 1 light chain mRNAs were constructed as previously described⁸ using the DNA insert from pAB λ 1-1 as probe¹³. A preliminary characterization of several λ 1-containing cDNA clones by restriction enzyme cleavage suggested the presence of somatic mutation in the S43-derived sequence. Restriction enzyme digests of three B1-8 clones were consistent with the sites predicted from the germ-line λ 1 DNA sequence^{2,13} but the S43-derived clone, pAB λ 1-15, had an additional *Kpn*I site in the J region (Fig. 1).

To precisely identify differences from the germ-line sequence, the longest cDNA clones pAB λ 1-9 (B1-8) and pAB λ 1-15 (S43), were sequenced. The sequence of the pAB λ 1-9 insert indicated that the clone began with codon 47 and extended in the 3' direction through the poly(A) (Fig. 1). This nucleotide

sequence was identical to that found in the BALB/c germ line. Peptide mapping of the B1-8 λ 1 chain also suggested that this chain contained the germ line-encoded sequence (T.I., unpublished).

The DNA sequence of the longest S43-derived clone, pAB λ 1-15, encoding the leader, V, J and most of the C region (up to the *Pvu*II site at codon 195) is presented in Fig. 2. (The rest of the λ 1 sequence in Fig. 2 was derived from pAB λ 1-9.) The S43 V-region sequence contained three somatic mutations when compared with the germ-line sequence; two in the $V_{\lambda 1}$ region at codons 80 and 87 and one in the J segment at codon 103. Otherwise the DNA sequence consisting of the leader, V, J and the available C region was exactly the same as that found in the B1-8-derived clone, pAB λ 1-9, or in the BALB/c λ germ-line DNA sequences¹³. The somatic mutations in the λ chain of S43 all occurred outside the CDRs—they are the first somatic mutations to be located outside these regions in a productive λ 1 gene. The mutations in codons 80 and 87 result in fairly conservative amino acid replacements; the new methionine (Met) residue at codon 87 is also found at that position in the λ 2 V region.

Comparison of the C57BL/6 and BALB/c λ 1 germ-line cDNA sequences beginning with the initiator Met codon and extending 3' to the poly(A) sequence showed complete identity. Given the polymorphisms evident in comparisons of heavy chain constant regions between these two strains¹⁵, the complete absence of polymorphic differences in these λ 1 sequences is notable.

To characterize somatic mutations present in the MOPC-315 λ 2 chain, the DNA sequence of p λ -1 was determined¹³ (Fig. 2). Six somatic mutations in the V region were found, at codons 38, 50, 94, 95 and 96. This sequence corrects the published amino acid sequence at positions 55 (where MOPC-315 λ 2 has the germ-line sequence), and at 145 and 146 and it removes the ambiguity at codons 147-148^{16,17}. All of the observed somatic mutations occurred in CDRs except for those found in codon 38. The precise junction between the $V_{\lambda 2}$ and $J_{\lambda 2}$ segments can be defined because the $J_{\lambda 2}$ germ-line sequence has recently been determined^{18,19}. The $J_{\lambda 2}$ sequence in p λ 2-1 is identical to the germ-line sequence.

These analyses have shown that somatic mutations in λ chains need not be localized in the CDRs. The sequence of an aberrantly rearranged λ 1 gene in MOPC-315 also showed somatic variation outside of the CDRs²⁰. It therefore seems likely that for λ , as for κ ³⁻⁷ and V_H genes⁸⁻¹², somatic mutation occurs throughout the variable region sequence.

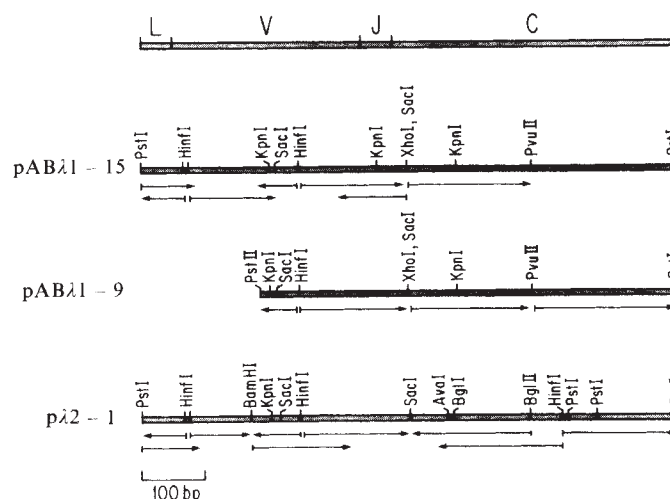
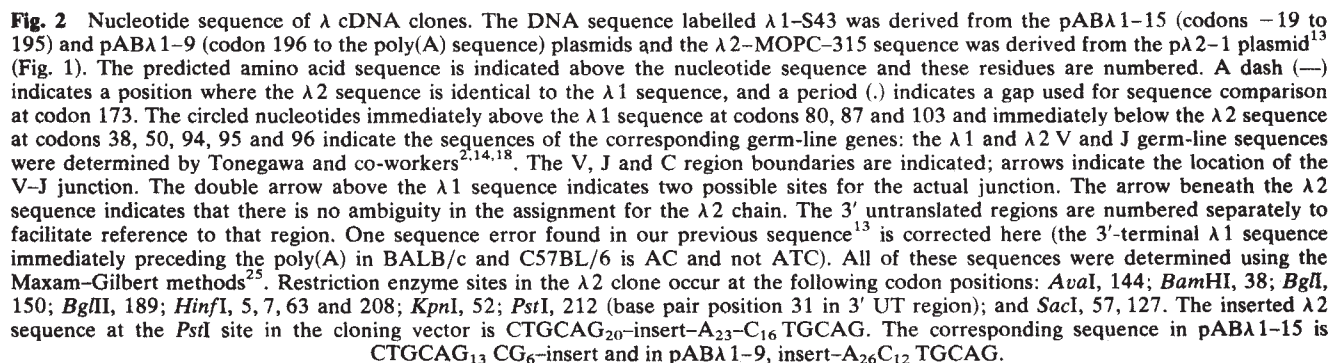


Fig. 1 Restriction map and DNA sequencing of λ cDNA clones. The clone pAB λ 1-15 was derived from the S43 hybridoma, pAB λ 1-9 from B1-8 and p λ 2-1 from the MOPC-315 myeloma¹³. The DNA sequence was determined using the Maxam-Gilbert²⁵ procedure on 5'-kinased DNA. The arrows indicate the DNA sequences determined in the 5' to 3' direction.



In previous experiments, all of the observed somatic variation in myeloma $\lambda 1$ chains was restricted to the CDRs¹. Furthermore, the existing data on V_{κ} and V_H shows a strong bias for somatic variation (as well as genetic variation) to occur in the CDRs (refs 8, 12 and A.B., M.P. & D.B., unpublished). There are two alternative possibilities for this clustering of somatic events: either there could be a predisposition for mutations in the CDRs, or selection could be operating. Preferential somatic mutation in the CDRs would require a remarkably specific mechanism but may occur. Selection could either act negatively, causing a loss of variants in the framework regions, or could act positively and favour proteins with variation in the CDRs. Negative selection seems unlikely because an excess of silent (synonymous) mutations in the framework compared to those from random mutational events would be expected. However the ratio between silent and replacement mutations in these λ sequences was approximately that expected from random mutation (25%): S43 and the MOPC-315 $\lambda 2$ had 1/3 and 1/6 silent mutations, respectively. Positive selection could be exerted either through better antigen binding due to variations in the CDRs or selective regulation by the immune system, but the expected bias in favour of replacement mutations might not be evident in the limited data available.

The constant regions of both the $\lambda 1$ light chain and $\gamma 2a$ heavy chain¹⁵ of the S43 hybridoma are germ-line DNA sequences although both variable regions show mutation⁸. Studies of the spread of somatic mutation away from the coding sequence suggest that barriers must exist to the spread of mutational activity. For κ light chains, no spread outside of the coding regions was observed⁵⁻⁷ although for $\lambda 1$, mutations just 3' to the J segments have been observed^{2,20}. Because the $V_{\lambda 1}$ gene is only 1,250 nucleotides from the $C_{\lambda 1}$ gene in the joined configuration², the lack of mutation in the S43C $\lambda 1$ region suggests that the barrier to mutational spread in the $\lambda 1$ gene is within the 1,250 nucleotides 5' to the $C_{\lambda 1}$ region. For heavy chains, extensive spread of mutation away from the rearranged V_H region has been observed^{9,10}.

There does not seem to be absolute specificity for mutations at particular nucleotides although similarities have been observed²¹. Interestingly, the three S43 $\lambda 1$ chain mutations occurred only at purine residues and we have noted a similar bias in the V_H region from the S43 hybridoma (9 out of 10)⁸. The six somatic mutations in the $\lambda 2$ cDNA, however, occurred at all four bases, three at purine and three at pyrimidine residues. The significance of a purine bias in some but not all data remains puzzling.

The MOPC-315 tumour expresses two λ mRNA species, one producing a normal $\lambda 2$ protein and the other producing a truncated $\lambda 1$ protein¹³. The V-J junction of the $\lambda 1$ sequence puts the reading of the rest of the mRNA out of frame²⁰. Presumably in the progenitor MOPC-315 cell, the $\lambda 1$ gene rearranged first but because a functional polypeptide could not be synthesized, the $\lambda 2$ gene was later rearranged. Following the productive $\lambda 2$ rearrangement, a similar level of somatic mutation occurred in both rearranged λ genes, the $\lambda 1$ gene having seven somatic mutations in its coding sequence²⁰ and the $\lambda 2$ containing six mutated bases (Fig. 2). Pech *et al.*⁵ found a related situation for κ genes where in one myeloma cell both an expressed and an unexpressed but rearranged κ gene showed similar levels of somatic variation.

Except for one report (HPC 16; ref. 22), data on somatic variation has revealed no variants appearing in association with the μ heavy chain isotype but many variants have occurred in association with other isotypes^{8,23}. This again was seen in the λ data: B1-8 has a μ heavy chain and no somatic variation in its λ chain, S43 has a $\gamma 2a$ heavy chain and somatic variation in both its heavy and λ light chain, and MOPC-315, with an α heavy chain, has somatic variation in its $\lambda 2$ light chain. Somatic mutation is not, however, obligatory during isotype switching: protein sequencing has identified germ-line sequences in non-IgM antibodies²³ as has nucleotide sequencing (A.B., M.P. & D.B., in preparation). It seems that isotype switching

may precede somatic mutation. Somatic mutation may be an inducible phenomenon (perhaps via T-cell help) in some stage of B-cell development after its encounter with antigen. More data will be needed to clarify these relationships.

Finally, these data allow a direct comparison of the nucleotide sequences of the $\lambda 1$ and $\lambda 2$ constant regions. The two genes are obviously related but differ at 30 out of 105 amino acids (Fig. 2). They are more different from one another than the three human λ sequences recently determined²⁴. They differ so extensively in their 3' flanking sequences that the homology is not very evident (and therefore in Fig. 2 the identities are not shown with dashes). The two genes probably diverged in a species ancestral to the present mouse species. By contrast, the human λ genes are closely related and therefore diverged quite recently, as did the mouse $\lambda 2$ and $\lambda 3$ genes as shown by protein sequence data¹⁷.

This work was supported by grants from the American Cancer Society and the NCI and a contribution from the Whitehead Charitable Foundation. A.B. was a postdoctoral fellow of the Medical Foundation. D.B. is an American Cancer Society Research Professor.

Received 1 March, accepted 18 May 1982.

1. Weigert, M. & Riblet, R. *Cold Spring Harb Symp. quant. Biol.* **41**, 837-846 (1976).
2. Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15**, 1133-1144 (1978).
3. Weigert, M., Gattmaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276**, 785-790 (1978).
4. Valbuena, O., Marcu, K. B., Weigert, M. & Perry, R. *Nature* **276**, 780-784 (1978).
5. Pech, M., Hochli, J., Schnell, H. & Zachau, H. G. *Nature* **291**, 668-670 (1981).
6. Selsing, E. & Storb, U. *Cell* **25**, 47-58 (1981).
7. Gershenfeld, H. K., Tsukamoto, A., Weissman, I. L. & Joho, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7674-7678 (1981).
8. Bothwell, A. L. M. *et al. Cell* **24**, 625-637 (1981).
9. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286**, 676-683 (1980).
10. Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. *Cell* **27**, 573-581 (1981).
11. Kataoka, T., Nikaido, T., Miyata, T., Moriwaki, K. & Honjo, T. *J. biol. Chem.* **257**, 277-285 (1982).
12. Crews, S., Griffen, J., Huang, H., Calame, K. & Hood, L. *Cell* **25**, 59-66 (1981).
13. Bothwell, A. L. M., *et al. Nature* **290**, 65-67 (1981).
14. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485-1489 (1978).
15. Schreier, P. H., Bothwell, A. L. M., Mueller-Hill, B. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4495-4499 (1981).
16. Dugan, E. S., Bradshaw, R. A., Simms, E. S. & Eisen, H. N. *Biochemistry* **12**, 5400-5416 (1973).
17. Azuma, T., Steiner, L. A. & Eisen, H. N. *Proc. natn. Acad. Sci. U.S.A.* **78**, 569-573 (1981).
18. Blomberg, B. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 530-533 (1982).
19. Miller, J., Selsing, E. & Storb, U. *Nature* **295**, 428-430 (1982).
20. Hozumi, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 7019-7023 (1981).
21. Baltimore, D. *Cell* **26**, 295-296 (1981).
22. Kocher, H. P., Berec, C., Schreier, M. H., Cosenza, H. & Jaton, J. C. *Eur. J. Immun.* **10**, 264-267 (1980).
23. Gearhart, P. J., Johnson, N. D., Douglas, R. & Hood, L. *Nature* **291**, 29-34 (1981).
24. Hieter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. *Nature* **294**, 536-540 (1981).
25. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).

Mapping class I gene sequences in the major histocompatibility complex

Larry R. Pease, Stanley G. Nathenson
& Leslie A. Leinwand

Department of Microbiology and Immunology and Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

The class I glycoproteins of the major histocompatibility complex (MHC) are products of closely linked genes on the 17th chromosome of the mouse. Four highly homologous cell surface proteins, K, D, L, and Qa-2 (refs 1-4), belonging to this multigene family have been analysed. Several additional members⁵⁻⁷, M, R, TL, are indicated by biochemical data and recent serological data suggest the family is larger⁸. Recent reports using recombinant DNA technology suggest that the class I system is an extensive multigene family⁹⁻¹². In this report we describe results of experiments using restriction enzyme digestion of DNA from a group of standard, congenic and MHC

recombinant mouse strains followed by hybridization to H-2 cDNA probes. Particular bands of hybridization can be assigned to specific H-2 haplotypes and, therefore, to the MHC region of the 17th chromosome. A number of the MHC polymorphic bands have been assigned to genetic map positions in the K and D-L regions of the H-2 complex. Additional class I gene sequences map outside the boundaries of the traditional H-2 complex (K, I, S, D), into the Qa-T1a region. At least one sequence is centromeric to the K locus. Further, amino terminal and carboxy terminal probes show different restriction patterns, suggesting that the exons corresponding to the different domains of the class I glycoproteins may have different evolutionary histories.

We used two cDNA probes⁹ to identify the DNA sequences bearing homology to the class I antigens. The first, pH-2.IIa, is homologous to the exons encoding from amino acid 167 through the carboxy terminus of the H-2K^b glycoprotein. The second probe, pH-2.III, is homologous to exons encoding from amino acid 63 to 160.

To determine whether the bands that hybridize to the cDNA probes are specific for any given haplotype and if these bands are located in the MHC, comparisons were made between hybridization patterns of class I cDNAs to genomic DNAs of standard mouse strains and those bearing (1) different MHC haplotypes but identical non-MHC backgrounds, and (2) identical MHC haplotypes but different non-MHC backgrounds. DNA from these standard and congenic mouse strains was digested with *Bam*HI endonuclease and examined by hybridization to the carboxy terminal cDNA probe pH-IIa (Fig. 1a). The results show that approximately 50% of the bands in the DNA of each haplotype are polymorphic when compared with

the other haplotypes (7/13 in the *b* haplotype, 9/16 in the *d* haplotype, 6/13 in the *k* haplotype). This polymorphism can be assigned to the MHC segment on the 17th chromosome because the congenic strains, BALB·B, BALB/c and BALB·K differ only at the MHC chromosomal segment. Schematic representation of the banding patterns and the distribution among three haplotypes is shown in Fig. 2. The degree of polymorphism observed by our analyses (approximately 50%) is probably an underestimate since dissimilar bands may be approximately equal in size and high molecular weight bands greater than 5 kilobases (kb) are not resolved as well as the smaller bands.

The distribution of *Bam*HI fragments homologous to the two probes was different (Fig 1a,b). Unlike the pattern of bands observed with the carboxy terminal probe H-2.IIa, the amino terminal probe, pH-2.III, distinguished between two groups of bands. One family consisted of four prominent bands which hybridized strongly to the probe while the remaining 7 to 9 bands hybridized weakly.

The availability of congenic mouse strains as well as recombinant mouse strains, that were derived from previously defined arrays of class I alleles, provided the means to assign the class I hybridizing bands not only to the MHC of chromosome 17 but also to specific map positions within the MHC. The results of a series of experiments (as in Fig. 1c) utilizing *b*, *d*, and *k* haplotypes and recombinant haplotypes, *a*, *h4*, *i3*, *i5*, *g2* and *g* are shown in Fig. 3. The mapping patterns fall into several groups. Bands 1, 3, 4, 5, 6, 7 and 17 cannot be mapped to any specific region within the MHC since these segments of the genome do not contain detectably polymorphic *Bam*HI sites. Two *Bam*HI bands, 2 and 18, map to the K end of the H-2 complex. Band 2 is present in the *d* haplotype, also present in the *g2* and *g* haplotypes, but absent in the *b* haplotype. Since *g2* and *g* are haplotypes derived from recombination events that substituted a portion of the *b* haplotype at the D end of the H-2 complex, band 2 must map to the region containing the alleles derived from the *d* haplotype, namely the K end. Band 18 can also be mapped to the K end since it is present in the *b* haplotype but is absent in the *d*, *g2* and *g* haplotypes.

Another group of bands appears to map to the D end of the H-2 complex. Band 12, for example, is present in the *d* and *a* haplotypes but not in the *b* haplotype. Recombinants *i3* and *i5* contain band 12 and have a haplotype-specific genetic sequence from I-C through the T1a region. Thus the presence of band 12 in these strains indicates that the location of the DNA fragments contained in band 12 must be at the D end. The absence of band 12 from the recombinant *h4* which does not have either the *d* or *a* haplotype sequences in the D region further supports this mapping assignment. Similar analyses can be used for the mapping of the other bands, 10, 11, 13, 14, 16, 20 and 22 to the D end.

The remaining bands, 8, 9, 15, 19 and 21 cannot be directly mapped by the recombinants tested. The ambiguity about the location of these bands arises because the genetic regions homologous to class I cDNA probes may map to the centromeric as well as the distal side of the MHC because the boundaries defined by the congenic and recombinant strains have not been established. However, it is possible to eliminate some regions of the MHC, where these bands cannot be located. For example, bands 15 and 19 which are present in the *a* haplotype but not the *k* haplotype must be absent from the K or I-A region of the H-2 complex. Similarly, bands 8 and 21, which are present in the *k* haplotype but in none of the other strains examined, should not be present in the K or Ia region. Band 9 is present in the *a* haplotype but absent from both *d* and *k* haplotypes and thus must map outside of the traditionally defined H-2 complex since the *a* haplotype contains both *d* and *k* alleles throughout the complex.

The genetic sequences mapping to the D-end of the H-2 complex have been analysed with additional restriction enzymes and a Qa-T1a congenic mouse line, B6.T1a^a (data not shown). The recombinant MHC haplotype of this inbred strain includes the K, D and L alleles of the *b* haplotype and the Qa-T1a

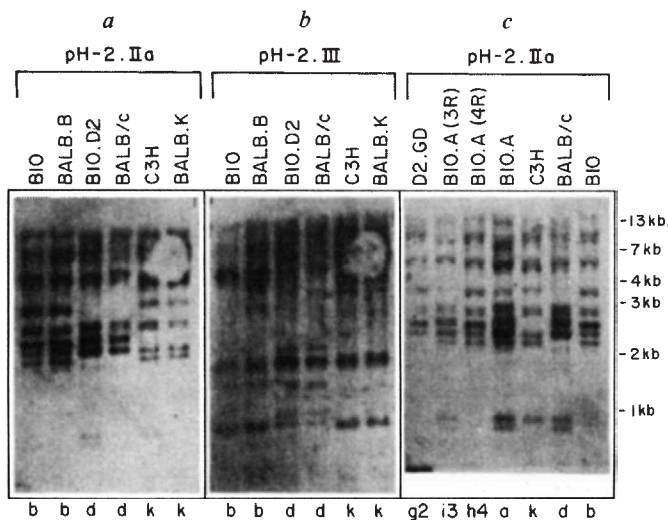


Fig. 1 Characterization of *Bam*HI genomic restriction fragments by hybridization to the cDNA probes, pH-2.IIa and pH-2.III. Splenic DNA was isolated from inbred strains maintained by Dr F. Lilly. Spleen cells were disrupted in 10 mM Tris (pH 7.4), 0.05 M EDTA, 0.5% SDS and 100 g ml⁻¹ proteinase K and incubated at 37 °C for 16 h. Protein contaminants were extracted from high molecular weight DNAs with phenol, chloroform, isoamyl alcohol, and the DNAs were recovered by ethanol precipitation. Samples of 12 µg of DNA from each strain were digested to completion with *Bam*HI and electrophoresed overnight on 0.8% Agarose gels in G-M buffer (0.18 M Tris, 0.18 M boric acid, 6 mM EDTA, pH 8.3) at 60 V. The DNA in the gels was transferred to nitrocellulose paper by the method of Southern¹³. Following transfer, the filters were prehybridized (5 × SSC, 50 mM PO₄, pH 7.4, 50 µg ml⁻¹ salmon sperm DNA, 50 µg ml⁻¹ poly(A), 1X Denhardt's solution) at 65 °C for 6 h. Radioactively labelled cDNA clones (2 × 10⁵ c.p.m. ml⁻¹) were added to prehybridization solution containing 20% dextran sulphate. After hybridizing for 6 h at 65 °C the filters were washed three times in 2 × SSC, 0.2% SDS at 65 °C and twice in 0.1 × SSC, 0.2% SDS at 55 °C for 20 min. The filters were exposed to X-ray film with an intensifying screen at -70 °C.

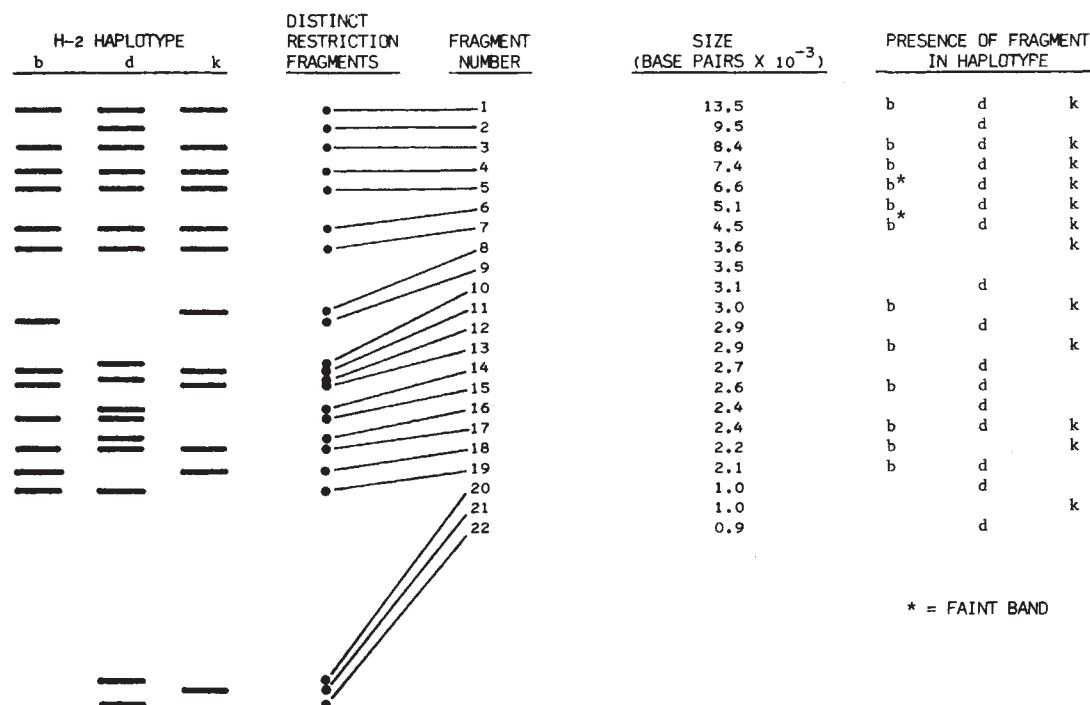


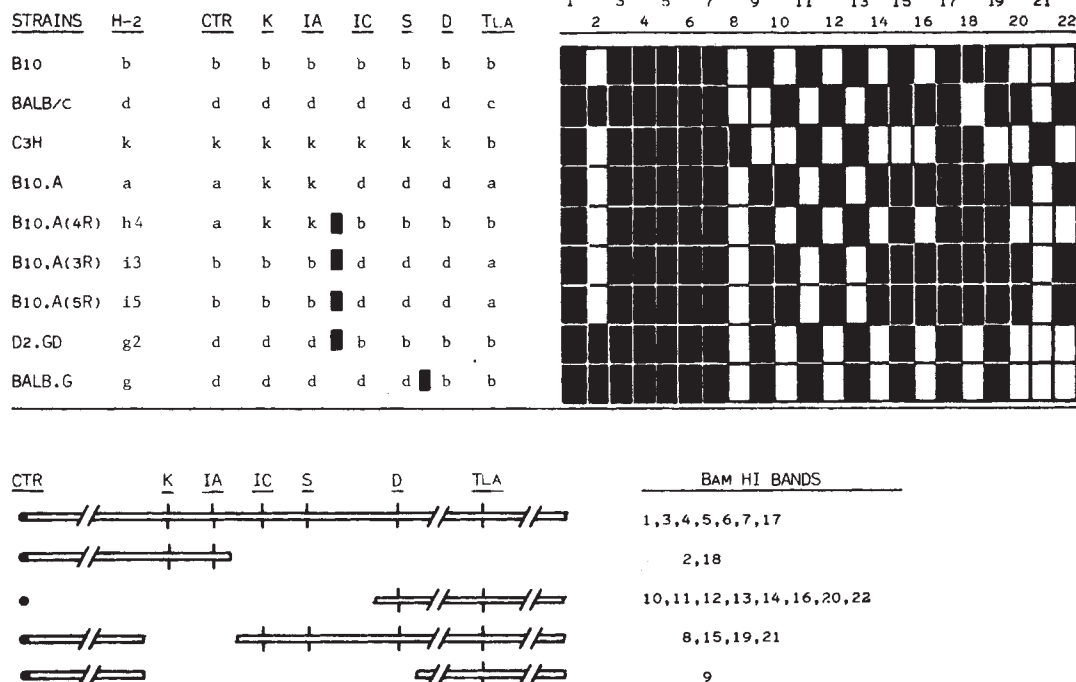
Fig. 2 Schematic representation of *Bam*HI fragments hybridizing to the carboxy terminal probe pH-2.IIa. This summary is based on several analyses, including those shown in Fig. 1. Bands were assigned a number according to their relative migratory position on 0.8% agarose gels in G-M buffer (see Fig. 1). Bands not distinguishably different among haplotypes were assigned the same number. Size estimates are based on band mobility relative to *Hind*-III fragments of λ phage DNA.

alleles of the *a* haplotype. A comparison of the banding patterns of B6.T1a^a with those from the strains B10(H-2^b), B10.A (H-2^a), and B10.A(3R) (H-2ⁱ³) (Fig. 3), subdivides the polymorphic D-end sequences into two groups, those mapping to the D-L region and those mapping to the Qa-T1a region. Similar mapping assignments of Class I genes to the Qa-T1a region have been reported by others^{11,12}.

A similar analysis has been carried out using the amino terminal probe pH-2.III (data not shown). Results are similar to the pattern seen with pH-2.IIa with the exceptions that two instead of one *d* haplotype specific fragment maps to the K end, and one fragment shared by the *d* and *a* haplotypes maps between H-2K and the centromere.

The different hybridization patterns observed with the amino terminal and carboxy terminal probes for the *Bam*HI fragments (Fig. 1), raise questions about the evolutionary relationships among the exons representing the protein domains of the class I antigens. For example, recent evidence suggests that the third protein domain of H-2K^b, (approximately amino acids 180 to 270), is critical for association with β_2 -microglobulin (K. Yokoyama and S.G.N., unpublished data). This region is quite important for the function of the molecule. Since other class I glycoproteins are believed to associate with β_2 -microglobulin¹⁵, they may also retain a similarly conserved domain. This domain is homologous to the carboxy terminal probe (pH-2.IIa), but not to the amino terminal probe, a relationship which could be

Fig. 3 Summary of H-2 recombinant *Bam*HI banding patterns. As described in Fig. 1, DNAs from the inbred strains B10, BALB/c, and C3H were compared with the DNA from the presumed B10.A, the *a* × *b* haplotype recombinants B10.A(3R), B10.A(4R), B10.A(5R), and the *b* × *d* haplotype recombinants D2.GD and BALB.G. The approximate positions of recombinant sites, marked with a bar, in all recombinant strains have been established by classical mapping using serological markers¹⁴. The presence or absence of different pH-2.IIa homologous *Bam*HI bands, as defined in Fig. 2, are shown (blank versus filled space) next to each strain examined. Unique bands resulting from recombination were not observed. The relative position of bands on the 17th chromosome is presented schematically.



responsible for the extensive strong cross-hybridization of the probe with numerous class I sequences throughout the MHC. In contrast, the amino terminal probe hybridized to restriction fragments, with various intensities, suggesting that those sequences homologous to the amino terminal portion of the molecules have evolved more rapidly than those homologous to the C-terminal portion.

The findings reported in this study indicate that a multigene family, homologous to genes encoding class I H-2 antigens, extends from the centromeric side of the K region through the Qa-T1a region on chromosome 17. Most of the polymorphic bands map the gene sequences distal, with respect to the centromere, to the complement components (S region) of the H-2 complex. Many of these sequences clearly are associated with the Qa and T1a regions. These data suggest that the H-2 complex is only a small part of a larger set of genes comprising the major histocompatibility complex.

We thank Diane McGovern for technical assistance and Catherine Whelan and Angela Stockhausen for secretarial help. This work was supported by NIH grants AI-07289 and AI-10702 (S.G.N.), GM-29020 (L.A.L.), and ACS grants IM-236 (S.G.N.) and NP-348 (L.A.L.).

Received 12 February; accepted 6 May 1982.

1. Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J. & Coligan, J. E. *A. Rev. Biochem.* **50**, 1025-1052 (1981).
2. Nairn, R., Nathenson, S. G. & Coligan, J. E. *Eur. J. Immun.* **7**, 495-503 (1980).
3. Coligan, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1134-1138 (1980).
4. Soloski, M. J., Uhr, J. W., Flaherty, L. & Vitetta, E. S. *J. exp. Med.* **153**, 1080-1093 (1981).
5. Ivanyi, D. & Demant, P. *Immunogenetics* **8**, 539-550 (1979).
6. Hansen, T. H. *et al. J. Immun.* **126**, 1713-1716 (1981).
7. Yokoyama, K., Stockert, E., Old, L. J. & Nathenson, S. G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7078-7082 (1981).
8. Ivanyi, D. & Demant, P. *Immunogenetics* **12**, 397-408 (1981).
9. Steinmetz, M. *et al. Cell* **24**, 125-134 (1981).
10. Cami, B., Bregegere, F., Ahastado, J. P. & Kourilsky, P. *Nature* **291**, 673-675 (1981).
11. Steinmetz, M. *et al. Cell* **25**, 683-692 (1981).
12. Margulies, D. H., Evans, G. A., Flaherty, L. & Seidman, J. G. *Nature*, **295**, 168-170 (1982).
13. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
14. Altman, P. L. & Kutz, D. D. in *Inbred and Genetically Defined Strains of Laboratory Animals: Part 1*, 125-127 (Federation of American Societies for Experimental Biology, Bethesda, 1979).
15. Vitetta, E. S. & Capra, J. D. *Adv. Immun.* **26**, 147-193 (1979).

Reactivation of latent feline leukaemia virus infection

Jennifer L. Rojko*, Edward A. Hoover*†, Sandra L. Quackenbush* & Richard G. Olsen*‡

* Department of Veterinary Pathobiology, † Comprehensive Cancer Center and ‡ Department of Microbiology, The Ohio State University, 484 W. 12th Avenue, Columbus, Ohio 43210, USA

In most cats exposed to the contagious¹ feline leukemia virus (FeLV), viral replication is contained in target haematopoietic tissues² and elicits humoral immunity to FeLV and to the feline oncornavirus-associated cell membrane antigen (FOCMA)³⁻⁶. Recently, we and others have considered that these ostensibly self-limiting infections might be persistent nonproductive (latent) infections in certain haematopoietic cells. This hypothesis could account for the relapsing viraemias⁷, protracted incubation periods⁸, persistently high titres of antiviral and anti-FOCMA antibodies⁸, appearance of FeLV p27 antigen in serum of otherwise non-viraemic animals⁹ and occurrence of FeLV-negative but FOCMA-positive leukaemias in naturally infected pet cats^{10,11}. Here we describe the reactivation of latent FeLV from myelomonocytic and lymphoid cells of cats immune to FeLV, cats bearing FeLV-negative tumours, and kittens congenitally exposed to FeLV. Furthermore, the reappearance of FeLV infection is suppressed by the host's immune system but this can be altered by adrenal corticosteroid hormones *in vivo* and *in vitro*.

Self-limiting (regressive) FeLV infection was established in 10 adult specific pathogen-free cats by two oral-nasal exposures to 10⁵ focus-forming units (FFU) of FeLV-R (Rickard) strain¹²

and confirmed by demonstration of transient group specific antigen (GSA) in peripheral blood mononuclear leukocytes (PBML)^{2,13}, 1-3 weeks after exposure and persistent FOCMA-L^{5,6,14} and FOCMA-S¹⁴ antibody responses 2-4 weeks after exposure.

At 16-20 weeks after exposure, femoral bone marrow cells, mesenteric lymph node cells, peritoneal macrophages and PBML were collected. Freshly isolated cells were negative for GSA by fixed-cell immunofluorescence assay¹³ and were negative for replicating FeLV by a viral infectivity assay¹⁵ (Table 1). However, high titres of FeLV were reactivated following isolation and *in vitro* propagation of marrow from 8 of 10 FeLV-immune cats but not from any of 4 FeLV-naive cats (Table 1). Identification of the reactivated agent from four cats as FeLV was confirmed by neutralization¹⁶ of its infectivity for clone 81 cells. This was achieved by preincubation with a 1:10 dilution of goat anti-FeLV serum or a 1:4 dilution of autochthonous cat anti-FeLV serum but not FeLV-naive goat or cat serum. Moreover, immunofluorescent examination of methanol-fixed, cytocentrifuged preparations of marrow cells demonstrated small clusters of GSA-positive proliferating myelomonocytic precursor cells (20% of total cells present). Neither the differentiated macrophage majority nor the mature granulocyte minority also present contained GSA. Refractoriness to retroviral replication, therefore, accompanies differentiation of feline myelomonocytic progenitor cells.

Stimulation of lymphocytes from lymph node but not bone marrow, by the feline T_h cell mitogen staphylococcal Protein A (Protein A)¹⁷ gave rise to low titres of recrudescence FeLV. Unstimulated nodal cells and PBML did not generate viral infectivity, nor was it found in cells activated by other T cell mitogens (phytohaemagglutinin, pokeweed mitogen, concanavalin A) or the B cell mitogen lipopolysaccharide¹⁸. Lymphocyte and macrophage growth factors and the tumour promoter, 12-O-tetradecanoyl-phorbol-acetate, also failed to reactivate GSA or viral infectivity in node cells or PBML. Thioglycollate-induced peritoneal macrophages did not reactivate GSA or viral infectivity (Table 1). Parallel cultures of marrow, lymphocytes and macrophages from FeLV-naive cats were negative (Table 1; ref. 19).

Bone marrow, tumour biopsies, PBML and serum were obtained from two pet cats referred to Ohio State University Veterinary Teaching Hospital for FeLV-negative lymphosarcoma of the mesenteric or popliteal lymph nodes (Table 1). In both cats, virus was reactivated from cultured marrow cells but not blood or tumour cells.

Finally, matings between two regressor females (cats 748B1, 1586) and two regressor males (cats 1604, 748B2) produced two litters of five kittens (kittens 1586-1, 2, 3; 748B1-1, 2) with evidence of congenital FeLV infections (Table 1). Both females were re-exposed by a single intraperitoneal dose of FeLV during the first week of gestation but remained negative for viraemia of marrow origin^{2,13} throughout the 9-week gestation period. Freshly isolated fetal (1586-1, 45 days gestation) and neonatal (1586-2, 3; 748B1-1, 2) lymph node, thymus and maternal placenta (1586) were negative for GSA and viral infectivity. Infectious virus was expressed in cultured marrow from three kittens (1586-1, 2; 748B1-1), cultured thymus of two kittens (1586-1; 748B1-1) and freshly isolated, as well as cultured, liver haematopoietic cells of the fetal kitten (1586-1).

The presence of FeLV replication in fetal haematopoietic cells, but not maternal placenta, suggested that transplacental passage of FeLV was not prevented by maternal anti-FeLV responses. Other workers have shown that in cats the majority of maternal antibody was transferred in colostrum²⁰⁻²². Passive (colostral) transfer of maternal FeLV-neutralizing antibody²³ to suckling neonates is perhaps responsible for maintenance of the neonatal infection in its non-productive, but reactivatable, form.

Therefore, latent FeLV infections are characterized by absence of infectious virus in freshly isolated marrow, node, macrophage and PBML, and reactivation of viral infectivity

Table 1 Reactivation of latent FeLV infections

Cell source	Days in culture	Type of FeLV infection							
		FeLV-immune		FeLV-negative lymphosarcoma		Congenital		FeLV-naïve	
		No. of cats positive for VI	VI	No. of cats positive for VI	VI	No. of cats positive for VI	VI	No. of cats positive for VI	VI
Marrow	0	0/10	<5	0/2	<5	0/5	<5	0/4	<5
Marrow	6	8/10	11,794 (580-74,500)	2/2	1,565 (975-2,060)	3/5	90 (40-120)	0/4	<5
PBML	0	0/10	<5	0/2	<5	0/2	<5	0/6	<5
PBML-PA	7	0/10	<5	0/2	<5	0/2	<5	0/6	<5
Lymph node	0	0/5	<5	ND	ND	0/4	<5	0/6	<5
Lymph node-PA	7	5/5	155 (10-675)	ND	ND	0/4	<5	0/6	<5
Peritoneal macrophages	0	0/3	<5	ND	ND	ND	ND	0/4	<5
Peritoneal macrophages	7	0/3	<5	ND	ND	ND	ND	0/4	<5
Tumour	0	ND	ND	0/2	<5	ND	ND	ND	ND
Tumour	7	ND	ND	0/2	<5	ND	ND	ND	ND
Tumour	28	ND	ND	0/2	<5	ND	ND	ND	ND
Thymus	0	ND	ND	ND	ND	0/4	<5	0/4	<5
Thymus	6	ND	ND	ND	ND	2/4	77 (44-110)	0/4	<5

Cells were collected by aspiration (marrow, peritoneal macrophages), surgical biopsy (node, tumour, thymus), or venipuncture (peripheral blood) from cats anaesthetized with ketamine. Peritoneal macrophages were collected 4 days after intraperitoneal inoculation of thioglycollate. Partial purification of mononuclear cells from marrow and blood was by density separation on Ficoll-metrizoate gradients (specific gravity 1,077). Marrow and peritoneal macrophages cultured in McCoy's 5A with 20% (marrow) to 40% (macrophages) horse serum by modifications of the procedure of Dexter and Testa³⁴. PBML, lymph node neoplastic lymph node, and thymus were cultured in RPMI 1640 and 20% fetal calf serum¹⁹. A minimum of 10^6 freshly isolated or cultured cells in 1 ml were disrupted by two freeze-thaw cycles prior to viral infectivity (VI) assay²³; 0.2 ml of disrupted cell-associated inoculum and 0.2 ml of cell-free supernatant from each culture were each applied to 5×10^4 indicator clone 81 cells in 24 cluster dishes. VI data, expressed as FFU per 10^6 mononuclear cells, is the sum of cell-associated virus and cell-free virus for each cat positive for reactivated virus, averaged between the number of cats positive for reactivated virus from marrow versus node, etc. The ranges for the individual VI sums are indicated in brackets. ND, not determined.

from cultured marrow cells and a minor subset of nodal lymphocytes. These data suggest that self-limiting (regressive) FeLV infections are actually latent (nonproductive) infections of myelomonocytic precursor cells in the bone marrow and Protein A-reactive cells in the lymph nodes.

A similar coupling of virus latency *in vivo* with viral reactivation *in vitro* described for the putative human herpesviral oncogen, the Epstein-Barr Virus (EBV)^{24,25} has been attributed to the presence versus absence of a host immune response against virally infected cells. By analogy, while those cells expressing FeLV antigens would be eliminated in regressor cats, nonproductively infected cells would survive. Therefore, we investigated the role of host antiviral defenses in the maintenance of FeLV infections in their dormant phases. Regressor cats mounted both cell-dependent and antibody-dependent killing of autochthonous marrow cells with reactivated virus (Table 2) but did not kill freshly isolated marrow cells. Cells from FeLV-naïve cats had minimal cytotoxicity (<20%) against autochthonous freshly isolated or cultured marrow cells and peritoneal macrophages (data not shown). Regressor PBML (but not regressor sera) also were cytotoxic for autochthonous, non-transformed peritoneal macrophages (Table 2) negative for demonstrable FeLV production (Table 1); a finding that we do not understand, but is similar to the recognition of lymphocyte-defined membrane antigen on nonproductively infected B cells by killer cells in EBV-exposed people²⁶. The biological relevance of the killing of FeLV-infected marrow cells by immune PBML and sera is suggested by the correlation of decreases in circulating neutrophil and lymphocyte numbers with curtailment of virus replication and appearance of FOCMA antibody in regressively infected cats^{2,27}.

The early containment in haemolymphatic tissues and establishment of regressive versus progressive infection *in vivo* is corticosteroid (CS)-sensitive²⁸. *In vitro* studies have shown that

minority populations of feline macrophages and lymphocytes are susceptible to productive FeLV infection^{19,29} and that the permissiveness of macrophages but not lymphocytes to FeLV infection can be augmented by CS^{19,29}. To examine this further we investigated the effect of CS on latent FeLV infections *in vivo* and *in vitro*. Cats with established regressive FeLV infection were treated with 10 mg per kg methylprednisolone acetate twice weekly for 4 successive weeks. They developed productive FeLV infection in marrow and diminished antibody to FOCMA-L 1-2 weeks after initiation of CS treatment (Table 3). While 3 of 5 cats were able to contain reactivated FeLV, 2 of 5 cats remained persistently viraemic 135 days after initiation of treatment. Comparable non-CS-treated untreated regressors maintained for 2-3 years in similar environments have never developed productive infection⁴.

Reactivation of FeLV from latently infected marrow myelomonocytic precursor cells and from peritoneal macrophages but not from lymph node or PBML occurred *in vitro* following CS treatment (Table 4). Reactivation of FeLV was associated with induction of cytoplasmic GSA in 15-30% (mean 21%) of untreated regressor cat bone marrow cells and 20-32% (mean 25%) of CS-treated marrow cells. The CS-mediated 2-3-fold amplification of viral production, therefore, is perhaps due to increased viral replication in a fixed population (20-25%) of marrow cells that contain latent FeLV infection.

Both lymphocytes and macrophages are critical to the early restriction of reactivated FeLV infection. However, ultimately T cells of marrow origin which migrate to the thymus are transformed^{2,30}. The relationships between the target cells for latent infection (myelomonocytic and T_H cells), acute infection (B cells and myelomonocytic cells)^{2,19} and transformation (T cells)³⁰ are as yet undetermined.

Our findings suggest that most cats infected by FeLV develop persistent viral infection of marrow cells. A minority of cats

Table 2 Cytotoxic immune responses of FeLV regressor cats with reactivated infections

Type of MCT	No. of cats positive	Microcytotoxicity (MCT) target		Peritoneal macrophages	
		Marrow	% Killing (Range)	No. of cats positive	% Killing (Range)
ADCC	6/6	89	(52-99)	2/5	46 (25-60)
CTL (PBML)	4/6	99	(84-99)	3/5	60 (59-61)
CML-regressor serum	5/6	91	(65-99)	0/5	<20
CML-reference serum	5/6	72	(41-99)	0/5	<20
CML-pre-exposure serum	0/3	<20		0/3	<20
ΔH -regressor serum	4/6	88	(47-99)	ND	ND
ΔH -reference serum	1/3	38	(38)	ND	ND
ΔH -pre-exposure serum	0/3	<20		0/3	<20
Media control	0/6	0		0/5	0

500 marrow cells or peritoneal macrophages from regressor cats were seeded in triplicate in 96-well cluster dishes and cultured as described in Table 1 to permit *in vitro* FeLV reactivation. On day 6, nonadherent cells were aspirated and the adherent cells washed twice with fresh media before the addition of autochthonous cells, serum or reference serum. Serum were diluted 1:4 in media and added in 0.05-ml aliquots. For antibody-dependent cell-mediated cytotoxicity (ADCC), washed adherent cells were incubated with heat-inactivated (ΔH , 56 °C, 30 min) autochthonous serum for 30 min at 37 °C, washed twice and incubated with plastic nonadherent PBML at an effector:target (E:T) ratio of 50:1. For cytotoxic T lymphocytes (CTL), autochthonous plastic nonadherent BML only were added at an E:T ratio of 50:1. For complement-mediated lysis (CML), fresh or fresh-frozen autochthonous serum (regressor or pre-exposure serum) was added. Alternatively, ΔH reference goat anti-FeLV serum and 0.05 ml of fresh FeLV-naive cat complement³⁵ diluted 1:4 in media were added. To evaluate the complement-independent, cytostatic potential of regressor serum alone, ΔH regressor, reference, or pre-exposure serum was added. Cytotoxicity media was RPMI 1640-20% fetal calf serum. Plates were incubated for 20 h at 37 °C to allow lysis by cat complement³⁵ and nonadherent cells were removed by washing. Plates were stained with Wright Giemsa and examined for morphological evidence of target cell necrosis. % Killing is the arithmetic mean of triplicate wells for each of the no. of cats positive for MCT: (>30% killing of variable versus media control was significant at $P < 0.05$ by Student's paired *t*-test); and

$$\% \text{ Killing} = 100 - \frac{\text{No. of viable adherent cells in variable well}}{\text{No. of viable adherent cells in media control well}}$$

Table 3 Effect of adrenal corticosteroids on latent FeLV infections *in vivo*

Days post treatment	No. of cats positive for GSA in:		Marrow VI		FOCMA-L	
	Marrow	Blood	No. positive	Mean Titre (Range)	Mean Antibody (Range)	
0	0/5	0/5	0/5	<5	71	(32-256)
7-14	1/5	1/5	4/4	22 (15-30)	9	(<2-32)
28	2/5	2/5	2/5	30 (25-35)	4	(<2-8)
63-135	2/5	2/5	2/5	732 (200-1175)	32	(16-64)

Five adult regressors were treated with 10 mg per kg methylprednisolone acetate (Depo-Medrol, Upjohn) at days 0, 4, 7, 11, 14, 18, 21. Mononuclear and polymorphonuclear cells in blood and marrow smears were examined for FeLV GSA by fixed cell immunofluorescence¹³. A minimum of 10^6 freshly isolated marrow cells were disrupted by two freeze-thaw cycles and assayed for VI¹⁵. VI is expressed as mean FFU per 10^6 cells. Antibody to FOCMA was determined by indirect membrane immunofluorescence using FL-74 target cells^{5,6} and is expressed as geometric mean titre.

Table 4 Effect of adrenal corticosteroids on reactivation of latent FeLV infections *in vitro*

Cell source	Days in culture	CS treatment	No. of cats positive	VI		% GSA-positive cells	
				Mean	(Range)	Mean	(Range)
Marrow	0	Absent	0/6	<5		0	
	6	Absent	8/10	11,794	(580-74,500)	21	(15-30)
	6	Present	9/10	30,500	(2,350-136,700)	25	(20-32)
Lymph node	0	Absent	0/10	<5		0	
	7	Absent	0/10	<5		0	
	7	Present	1/10	15		0	
PBML	0	Absent	0/10	<5		0	
	7	Absent	0/10	<5		0	
	7	Present	1/10	10		0	
Peritoneal macrophages	0	Absent	0/5	<5		0	
	7	Absent	0/5	<5		0	
	7	Present	3/5	517	(50-950)	ND	

Cells were obtained from adult regressors and cultured as described for Table 1. 10^{-6} M CS (hydrocortisone hemisuccinate, Solucroref, Abbot) was included as indicated. VI is expressed as FFU per 10^6 cells (freeze-thaw lysate); % GSA-positive cells is expressed as arithmetic mean.

experience highly expressed persistent infections of marrow haemolymphatic progenitor cells, lymphoid tissues and epithelia², with consequent virus excretion³¹, immunosuppression^{32,33} and subsequent FeLV-related proliferative or antiproliferative disease⁴. Most FeLV-exposed cats (>70%), however, develop transient productive infection of haemolymphatic cells which is curtailed by host immune responses^{2,5}. Nevertheless, FeLV persists as an integrated latent infection in marrow myelomonocytic precursor cells and (certain nodal) T lymphocytes and can be reactivated by liberation

from host immunological controls through culture *in vitro* or immunomodulation *in vivo*. Comparable degrees of viral recrudescence *in vivo* may occur in response to endogenous hormones, senescence, or other immunosuppressive or promoting influences. These observations provide both a progression and a link between host-retrovirus relationships in the cat and those species in which greater degrees of viral genomic latency predominate, such as the ecotropic retroviral infections of cattle, gibbon apes, wild mice and the oncogenic human, simian and avian herpesvirus infections.

Received 11 November 1981; accepted 5 May 1982.

1. Jarrett, W. F. H. *et al. J. natn. Cancer Inst.* **51**, 833-841 (1973).
2. Rojko, J. L., Hoover, E. A., Mathes, L. E., Olsen, R. G. & Schaller, J. P. *J. natn. Cancer Inst.* **63**, 759-768 (1979).
3. Sarma, P. S. & Log, T. *Virology* **54**, 160-169 (1973).
4. Hoover, E. A., Olsen, R. G., Hardy, W. D. Jr, Schaller, J. P. & Mathes, L. E. *J. natn. Cancer Inst.* **57**, 365-369 (1976).
5. Essex, M., Klein, G., Snyder, S. P. & Harrold, J. B. *Int. J. Cancer* **8**, 384-390 (1971).
6. Essex, M., Klein, G., Snyder, S. P. & Harrold, J. B. *Nature* **233**, 195-197 (1971).
7. Post, J. W. in *Feline Leukemia Virus* (eds Hardy, W. D., Essex, M. & McClelland, A. J.) 151-159 (Elsevier-North Holland, New York, 1980).
8. Essex, M., Hardy, W. D. Jr, Cotter, S. M., Jakowski, R. M. & Sliski, A. *Infect. Immunity* **11**, 470-475 (1975).
9. Lutz, H., Pedersen, N. C., Higgins, J., Harris, C. W. & Theilen, G. H. in *Feline Leukemia Virus* (eds Hardy, W. D., Essex, M. & McClelland, A. J.) 497-505 (Elsevier-North Holland, New York, 1980).
10. Hardy, W. D. Jr, Zuckerman, E. E., MacEwen, E. G., Hayes, A. A. & Essex, M. *Nature* **270**, 249-251 (1977).
11. Hardy, W. D. Jr. *et al. Cold Spring Harb. Conf. on Cell Proliferation* **7**, 677-699 (1980).
12. Rickard, C. G., Post, J. E., Noronha, F. & Barr, L. M. *J. natn. Cancer Inst.* **42**, 987-1014 (1969).
13. Hardy, W. D. Jr, Hirshaut, Y. & Hess, P. *Bibl. Haemat.* **39**, 778-799 (1973).
14. Sliski, A., Essex, M., Meyer, C. & Todaro, G. J. *Science* **196**, 1336-1339 (1977).
15. Fischinger, P. J., Blevins, C. S. & Nomura, S. *J. Virol.* **14**, 177-179 (1974).
16. Schaller, J. P. & Olsen, R. G. *Infect. Immunity* **12**, 1405-1410 (1975).
17. Rojko, J. L., Finn, B. L., Hoover, E. A. & Olsen, R. G. *Clin. exp. Immun.* (submitted).
18. Rojko, J. L., Hoover, E. A., Finn, B. L. & Olsen, R. G. *Int. Archs Allergy appl. Immun.* (in the press).
19. Rojko, J. L., Hoover, E. A., Finn, B. L. & Olsen, R. G. *J. natn. Cancer Inst.* **67**, 899-910 (1981).
20. Schultz, R. D., Scott, F. W., Duncan, J. R. & Gillespie, J. H. *Infect. Immunity* **9**, 391-393 (1974).
21. Causse-Vaills, C., Vérain, A. & Vérain, A. *C. r. hebdom. Séanc. Acad. Sci., Paris* **252**, 2453-2455 (1961).
22. Tomada, I. *Jap. J. vet. Sci.* **24**, 337-348 (1962).
23. Hoover, E. A., Schaller, J. P., Mathes, L. E. & Olsen, R. G. *Infect. Immunity* **16**, 54-59 (1977).
24. Rickinson, A. B., Jarvis, J. E., Crawford, D. H. & Epstein, M. A. *Int. J. Cancer* **14**, 704-715 (1974).
25. Rickinson, A. B., Crawford, D. H. & Epstein, M. A. *Clin. exp. Immun.* **28**, 72-79 (1977).
26. Svedmyr, E. & Jondal, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1622-1626 (1975).
27. Hoover, E. A., Rojko, J. L. & Olsen, R. G. *Cold Spring Harb. Conf. on Cell Proliferation* **7**, 635-651 (1980).
28. Rojko, J. L., Hoover, E. A., Mathes, L. E., Krakowka, S. & Olsen, R. G. *Cancer Res.* **39**, 3789-3791 (1979).
29. Hoover, E. A., Rojko, J. L., Wilson, P. L. & Olsen, R. G. *J. natn. Cancer Inst.* **67**, 889-898 (1981).
30. Cockerell, J. L., Krakowka, S., Hoover, E. A., Olsen, R. G. & Yohn, D. S. *J. natn. Cancer Inst.* **57**, 907-913 (1976).
31. Francis, D. P., Essex, M. & Hardy, W. D. Jr *Nature* **269**, 252-254 (1977).
32. Perryman, L. E., Hoover, E. A. & Yohn, D. S. *J. natn. Cancer Inst.* **49**, 1354-1362 (1972).
33. Mathes, L. E., Olsen, R. G., Hebebrand, L. C., Hoover, E. A. & Schaller, J. P. *Nature* **274**, 687-698 (1978).
34. Dexter, T. M. & Testa, N. G. *Meth. Cell Biol.* **14**, 387-405 (1976).
35. Grant, C. K., De Boer, D. J., Essex, M., Worley, M. B. & Higgins, J. J. *Immun.* **119**, 401-406 (1977).

Culture of macrophage cell lines from normal mouse bone marrow

Jaroslav Činátl, Eduard Paluska, Vladimír Chudomel, Věra Malásková & Milan Elleder*

Institute of Haematology and Blood Transfusion, 128 20 Prague 2, U nemocnice 1, Czechoslovakia

* 1st Department of Pathology, Faculty of Medicine, 128 00 Prague 2, Studničkova 2, Czechoslovakia

The use of semisolid medium for the culture and cloning of haematopoietic cells^{1,2} has helped our understanding of their proliferation and differentiation. It has been shown that mixed colonies of granulocytes and macrophages developed, in the presence of colony-stimulating factor (CSF)^{3,4}, from their common precursor, granulocyte-macrophage colony-forming cells (GM-CFC)⁵. Attempts at recloning these colonies in semisolid medium suggested that granulocytes and macrophages were differentiated cells incapable of further proliferation^{6,7}. However, our studies on cultures of larger numbers of cells⁸ demonstrate that while this may be the case for granulocytes, macrophages seem to be capable of long-term proliferation.

The bone marrow for culture of GM-CFC was obtained from ICR SPF mice (Velaz). We used the technique of Metcalf⁹ to establish the primary culture; MEM medium¹⁰ was supplemented with 0.2% Bactotryptose (Difco)¹¹, 20% heat-inactivated pooled human serum (HS) or native fetal calf serum (FCS) (VŠV), 10% lung conditioned medium (LCM)¹², and 0.34% Bacto-Agar (Difco). For the culture we used Müller flasks (Cavalier). The primary culture was established from 5×10^5 nuclear bone marrow cells in 5 ml of medium. After 7 days at 37 °C in an ambient atmosphere plus 5% CO₂, the cells were counted and then transferred to fresh agar medium at a minimum concentration of 5×10^4 per 5 ml. Further passages were done in the same conditions approximately every 15 days. The cells in semisolid agar medium were stained with aceto-orceine.

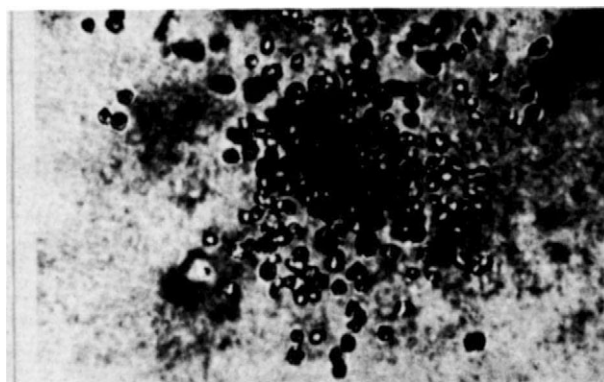


Fig. 1 A colony of macrophages from normal mouse bone marrow grown in semisolid medium; 15-day culture after the 22nd passage (330 days *in vitro*). Native preparation. $\times 63$.

Cells of the 20th passage were transferred to liquid medium for examination and 1 ml aliquots were put into Leighton test tubes with a cover glass. After 3 weeks of culture, colonies of adhering cells suitable for processing were formed on the cover glass.

Pinocytosis was studied by means of dextran sulphate. (molecular weight (MW) 5×10^5 ; Pharmacia); phagocytosis by means of latex particles (Imuna) and a suspension of autoclaved yeast. The receptors for IgG, complement, and IgM were determined by mouse antibodies to sheep erythrocytes. Lysozyme was demonstrated by anti-lysozymal serum using immunofluorescent techniques. The tests were performed according to Goud *et al.*¹³ and the peroxidase reaction was carried out according to Graham and Karnovsky. For the determination of nonspecific esterase α -naphthyl acetate was used which could be blocked by E-600 (3×10^{-5} M). For the determination of chloroacetyl esterase, naphthol-ASD-chloroacetate and acid phosphatase naphthol-ASBI-phosphate were used. The coupler was hexazonium-*p*-rosaniline throughout. The procedures are described by Lojda *et al.*¹⁴.

Ten macrophage cell lines have been obtained so far, the first of which is now in the 22nd passage (330 days *in vitro*) in

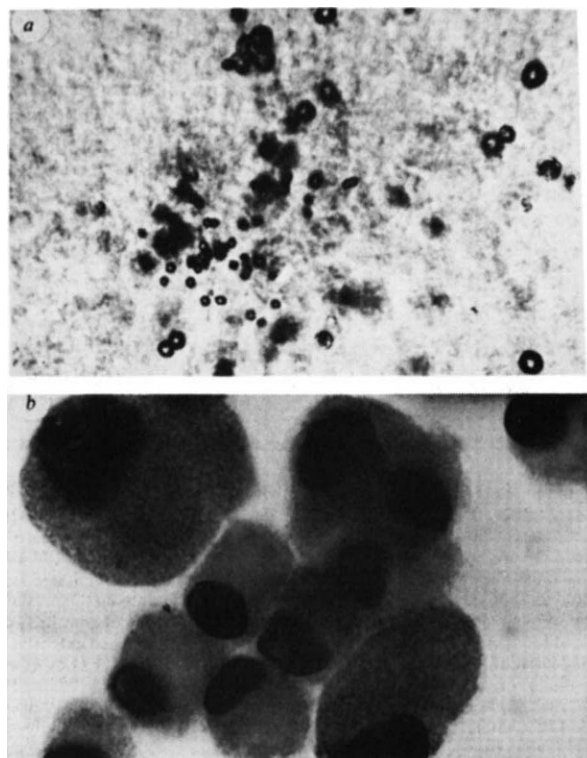


Fig. 2 *a*, A cluster of macrophages (starting colony) from a 7-day culture after the 15th passage in semisolid medium. The cluster consists of cells of various sizes. Small cells (10–14 μm in diameter) predominate in the culture. It is these cells which proliferate most readily are believed to maintain the continuous growth of the culture. Native preparation. $\times 30$. *b*, Small and large macrophages; 15-day culture after the 15th passage in semisolid medium. Fixed and stained with acetoorceine. $\times 630$.

medium with HS; the other three are in the 19th, 16th and 15th passage, and six lines are between the 5th and 10th passage. All of the latter nine lines were grown in FCS-containing medium. The number of seeded cells increased fivefold in ~ 15 days after every passage. The macrophages grew in colonies (>40 cells) (Fig. 1), and clusters (3–40 cells), with cells varying in size from approximately 10 μm to 30 μm (Fig. 2). Larger colonies normally consist of smaller cells and large macrophages are frequently found singly or in clusters. When stained in agar medium, both large and small cells show excentrically located nuclei with relatively rich cytoplasm (Fig. 2b).

Pinocytosis was found in 97.2% of cells, 95.8% of cells phagocytosed latex particles, and 93.6% phagocytosed yeast (Fig. 3). Receptors for IgG and complement were present in

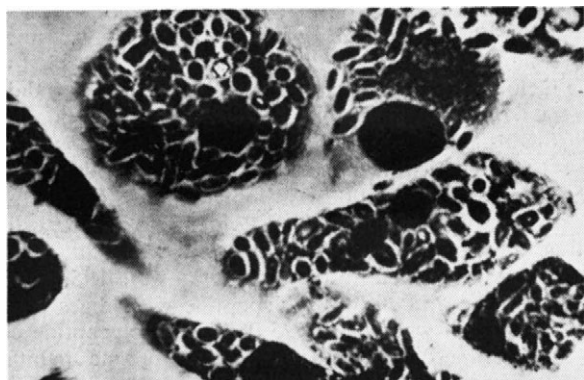


Fig. 3 Macrophage phagocytic activity after 20th passage. When transferred to liquid medium, phagocytes formed colonies of adhering cells on the surface of the cover glass. Oval forms within the cells are yeasts. Fixed with methanol. Phase contrast; $\times 360$.

89.3% and 86.1%, respectively but IgM-receptors could not be demonstrated. The immunofluorescent test for lysozyme was positive.

Peroxidase was negative in all cells. Nonspecific esterase was found in all cells, its activity being almost completely blocked by E600, and persisted only sporadically. Similarly, chloroacetyl esterase and acid phosphatase were positive in all cells. The activity of the hydrolases tested was high in most of the cells.

The assumption that the passaged cells are macrophages is supported by the demonstration of receptors for IgG and complement, and the absence of receptors for IgM¹³, and by the cytochemical examination, despite remarkably high chloroacetyl esterase activity.

Macrophages can be preserved in liquid medium for many weeks^{15,16}, but are believed to be incapable of passaging¹⁶. However, the use of semisolid media, which minimizes the number of cells adhering to the culture surface, allows long-term growth of macrophage populations from the original mixed granulocyte-macrophage colonies. The medium for long-term culture and passaging of macrophages must be supplemented with a suitable serum and LCM. The establishment of long-term cultures and passaging of macrophages from normal mouse bone marrow *in vitro* will be a useful tool for the further characterization of these very important cells.

Received 16 October 1981; accepted 14 May 1982.

1. Pluznik, D. H. & Sachs, L. *J. cell. Physiol.* **66**, 319–324 (1965).
2. Bradley, T. R. & Metcalf, D. *Aust. J. exp. biol. Sci.* **44**, 287–300 (1966).
3. Pluznik, D. H. & Sachs, L. *Expl Cell Res.* **43**, 553–563 (1966).
4. Metcalf, D. *J. cell. Physiol.* **76**, 89–100 (1970).
5. Metcalf, D. in *Hemopoietic Colonies*, 5–11 (Springer, Berlin 1977).
6. Metcalf, D., Johnson, G. R. & Mandel, T. E. *J. cell. Physiol.* **98**, 401–420 (1978).
7. Dexter, T. M., Garland, J., Scott, D., Solnick, E. & Metcalf, D. *J. exp. Med.* **152**, 1036–1047 (1980).
8. Činát, J., Paluska, E. & Chudomel, V. *Expl Hemat.* **9**, Suppl. 9, 113–113 (1981).
9. Metcalf, D. in *Hemopoietic Colonies*, 13–15 (Springer, Berlin, 1977).
10. Eagle H. *Science* **130**, 432–437 (1959).
11. Činát, J. *et al. Čs. fysiolog.* **29**, 33–38 (1980).
12. Sheridan, J. W. & Metcalf, D. *J. cell. Physiol.* **81**, 11–32 (1973).
13. Goud, T. J. L. M., Schotte, C. & van Furth, R. *J. exp. Med.* **142**, 1180–1199 (1975).
14. Loida, Z., Gossrau, R. & Schiebeler, T. H. *Enzyme Histochemistry* (Springer, Berlin, 1979).
15. Bennet, B. *Am. J. Path.* **48**, 165–181 (1966).
16. Adams, D. O. *Meth. Enzym.* **58**, 494–506 (1979).

Biosynthesis of membrane factor B by mouse peritoneal macrophages

Yuet M Ooi & Boon S. Ooi

Division of Nephrology, Department of Internal Medicine, University of Cincinnati Medical Center, Cincinnati, Ohio 45267, USA

The biosynthesis of many of the complement proteins by cells of the monocyte-macrophage series has been established¹. Studies on these cells using radiolabelled amino acids have demonstrated synthesis of precursor proteins (pro-complement)^{2–5}, and the native complement protein similar in size to that found in the plasma. However, synthesis of membrane complement proteins has not been demonstrated, although it has been suggested by previous studies using indirect techniques^{6–10}. In particular, there is evidence for the membrane-associated factor B in human lymphocytes^{11,12}. We report here that resident and thioglycollate-stimulated mouse peritoneal macrophages synthesized, in short-term primary cultures membrane factor B of molecular weight (MW) 95,000 and secreted factor B (MW 90,000) as single chain polypeptides. We also found a large single chain polypeptide with an approximate molecular weight of 195,000, which may be the putative factor B precursor.

Resident and thioglycollate-stimulated mouse (C57BL/6J) peritoneal macrophages were cultured at 37 °C for 18–48 h in Dulbecco's modified Eagle's medium (DMEM) in serum-free conditions in the presence of 60 mM 6- ϵ -amino caproic acid and radiolabelled with 35 S-methionine (150 $\mu\text{Ci ml}^{-1}$ of medium). At the end of the culture, the supernatants of the

Fig. 1 SDS (1%)–polyacrylamide (9%) gel electrophoresis (Laemmli)²³ of reduced and alkylated ³⁵S-methionine labelled intracellular factor B antigen from resident and thioglycollate-stimulated mouse peritoneal macrophages (mouse strain C57BL/6J). Resident and thioglycollate-stimulated cells were collected and cultured as described previously². Resident cells were plated at 13×10^6 and thioglycollate-stimulated cells at 10×10^6 cells on to 60×15 mm dishes in Dulbecco's modified Eagle's medium, 20% FCS (heated at 56 °C for 2 h). The cells were incubated in a humidified atmosphere of 95% air:5% CO₂ at 37 °C for 45 min to allow adherence and then non-adherent cells were rinsed off by several washes with Hank's balanced salt solution (HBSS). The adherent cells were cultured for 18 h in DMEM lacking cold methionine (incomplete DMEM) and containing 60 mM 6- ϵ -amino caproic acid (EACA) and ³⁵S-methionine ($150 \mu\text{Ci ml}^{-1}$ specific activity $>50 \text{ Ci mmol}^{-1}$; NEN). At the end of the culture, the medium was collected and the adherent cells were lysed with cold lysate buffer consisting of Tris-HCl (50 mM), KCl (100 mM), Triton X-100 (0.5%), deoxycholate (0.5%), EDTA (10 mM), EACA (100 mM), and phenylmethylsulphonyl fluoride (PMSF; 2 mM). Radiolabelled culture material was centrifuged at 10,000 r.p.m. in a microfuge at 4 °C for 45 min. To the extracellular culture medium was added EDTA (10 mM), PMSF (2 mM), and EACA (100 mM) final concentration and the extracellular culture medium was centrifuged at 150g for 10 min. The radiolabelled supernatants were immunoprecipitated with cross-reactive goat IgG antibody to human factor B (Atlantic Antibodies), added at two-fold antibody excess to yield 30 μg of immunoprecipitate, in the presence of ovalbumin with normal mouse serum as cold carrier². The control was mouse antiserum to ovalbumin reacted with ovalbumin in the presence of goat IgG; the antibody to ovalbumin was absorbed with mouse serum before use. After incubation for 3 h at room temperature, the immunoprecipitates were washed extensively. The precipitates were solubilized by boiling for 4 min in 80 μl of sample buffer pH 8.1 consisting of: SDS (1%), Tris-HCl (10 mM), Na₂-EDTA (1 mM) and β -mercaptoethanol (4%). After boiling, the sample was incubated with 120 mM iodoacetamide for 1 h at 37 °C and centrifuged at 10,000 r.p.m. for 2 min. The fluorograph following SDS-PAGE was developed by standard techniques²⁴ after fixing the slab gel in 7% glacial acetic acid with 50% methanol, followed by 1 h in Enhance (NEN) and 1 h in distilled water. The gel was dried under vacuum and exposed to a Kodak X-Omat-RP film at -70 °C for 7–10 days, depending on the radioactivity of the immunoprecipitates. Lane *a* shows mouse factor B antigen precipitated from cold carrier mouse serum stained with Coomassie blue, migrating as a protein of MW 90,000. The goat antibody is also detected: H, heavy chain; L, light chain. Lanes *b* to *i* represent the fluorograph of the SDS-PAGE of immunoprecipitated ³⁵S-methionine-labelled protein. Lane *h* shows newly synthesized ³⁵S-methionine-labelled reduced and alkylated intracellular factor B antigen from lysates of cultures of resident mouse peritoneal macrophages. Note the presence of the 195,000-MW and 95,000-MW bands compared with the control immunoprecipitate (*f*). Lane *i* shows these bands but is from intracellular factor B antigen from thioglycollate-stimulated mouse macrophages, which is not found in the control immunoprecipitate from stimulated cells (*g*). In *d* (resident cells) and *e* (thioglycollate-stimulated cells), reduced and alkylated extracellular factor B antigen is shown as a protein band of 90,000 MW, not detected in the controls (*b*, resident cells; *c*, thioglycollate-stimulated cells). ¹⁴C-labelled molecular weight markers are shown in *j*.

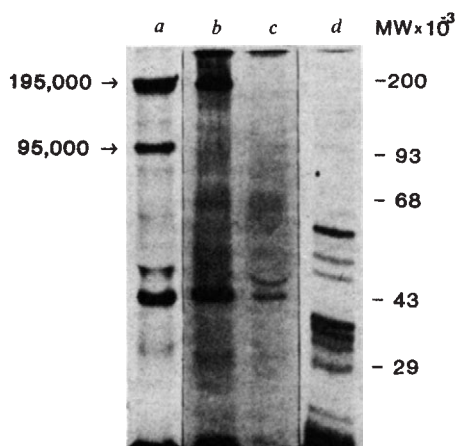
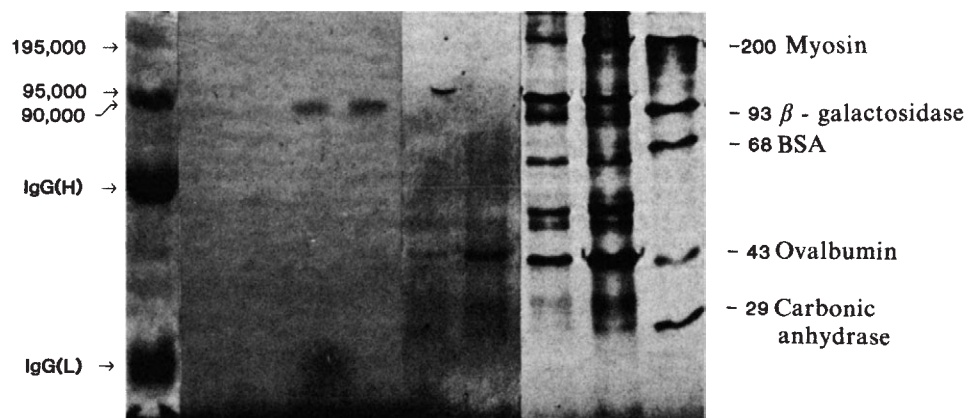


Fig. 2 For these experiments thioglycollate-stimulated and control cells cultured for 18 h in serum-free DMEM containing 60 mM ϵ -aminocaproic acid (EACA) and $150 \mu\text{Ci ml}^{-1}$ ³⁵S-methionine. The cells were rinsed and incubated with trypsin (tosylphenylethylchloromethylketone, TPCK) added at $10 \mu\text{g ml}^{-1}$ of HBBS at 37 °C for 45 min. At the end of the incubation, the cells were rinsed and lysed with cold lysate buffer with added protease inhibitors (EDTA, EACA and PMSF). A similar lysate was prepared without protease inhibitors and then incubated at 37 °C for 45 min with $10 \mu\text{g ml}^{-1}$ of trypsin. At the end of the incubation a twofold excess of soybean trypsin inhibitor (SBTI) was added. Immunoprecipitation and SDS-PAGE was then performed. Lane *b* shows the factor B precipitate of the ³⁵S-labelled thioglycollate-stimulated macrophages trypsinized off the cell surface. Note the disappearance of the 95,000-MW protein band and the presence of the 195,000-MW polypeptide chain. In lane *a* is the non-trypsinized control. Lane *c* is the control immunoprecipitation with anti-ovalbumin of a culture following trypsinization showing absence of the corresponding bands. Lane *d* shows precipitate of the lysate not treated with protease inhibitors. Note the disappearance of the 95,000-MW and the 95,000-MW polypeptides and the generation of several small peptide fragments. ¹⁴C-labelled molecular weight standards are shown on the right.

radiolabelled medium and of the cell lysates were immunoprecipitated with goat IgG antibody to human factor B, which is cross-reactive with mouse serum factor B. The antibody was shown by immunoelectrophoresis to be specific for factor B antigen of human and murine serum. Factor B antigen immunoprecipitated from mouse serum with the same antibody migrated as a single chain polypeptide of 90,000 MW (Fig. 1*a*) on SDS (1%) polyacrylamide gel (9%) electrophoresis (SDS-PAGE) of the immunoprecipitate. A similar SDS-PAGE of the immunoprecipitate obtained from ³⁵S-methionine labelled cell lysates of resident and thioglycollate-stimulated cells, detected several intracellular polypeptide chains and intermediate fragments: a large 195,000-MW polypeptide chain, a 95,000-MW chain, and several smaller MW polypeptide chains (88,000–64,000 and smaller) (Fig. 1*h* and *i*). The SDS-PAGE of the immunoprecipitate obtained from the extracellular culture yielded a single chain polypeptide of 90,000 MW (Fig. 1*d* and *e*) similar in size to that of mouse serum factor B.

³⁵S-methionine labelled thioglycollate-stimulated cells were incubated with trypsin ($10 \mu\text{g ml}^{-1}$) at 37 °C for 45 min, to remove membrane proteins, then rinsed and lysed. SDS-PAGE of the factor B immunoprecipitates of these cell lysates did not show the band representing the 95,000 MW polypeptide chain, suggesting its association with the membrane (Fig. 2*b*); control cells not incubated with trypsin retained their bands (Fig. 2*a*).

The intracellular 195,000-MW polypeptide was unaffected by trypsinization of the cell surface. Evidence for its intracellular location was obtained by experiments in which intracellular lysates were trypsinized; in these conditions the 195,000-MW protein and the 95,000-MW polypeptide were digested (Fig. 2*d*). To establish whether the 95,000-MW polypeptide chain was a membrane component, macrophage plasma membrane proteins were labelled by lactoperoxidase-catalysed radioiodination. The immunoprecipitate showed the presence of the 95,000-MW polypeptide (Fig. 3*b*) identical in size to the newly synthesized 95,000-MW polypeptide chain obtained from ³⁵S-methionine labelled cells (Fig. 3*a*). Macrophages cultured in

the presence of cycloheximide showed almost total abrogation of the expression of the 95,000-MW polypeptide chain (Fig. 3c) followed by its reappearance in cells radioiodinated after a 36 h culture in fresh medium following the removal of cycloheximide (Fig. 3e); these experiments therefore established *de novo* synthesis of this molecule and ruled out the possibility of absorption onto the cell membrane.

To further exclude absorption of membrane factor B studies were done using radiolabelled non-adherent peritoneal cells which did not show the presence of the 95,000-MW polypeptide chain (Fig. 3f). As a fourfold excess of radiolabelled non-adherent cells compared to the radioiodinated adherent cell population was examined, differential absorption of protein could not account for the different absorption patterns. Finally, it must be noted that on parallel SDS-PAGE analysis of intracellular, extracellular, and membrane labelled material, the extracellular (secreted) factor B has a different molecular size (MW 90,000) compared to the membrane and the cell lysate factor B (MW 95,000), which indicates that there are two distinct molecules one of which is a membrane component.

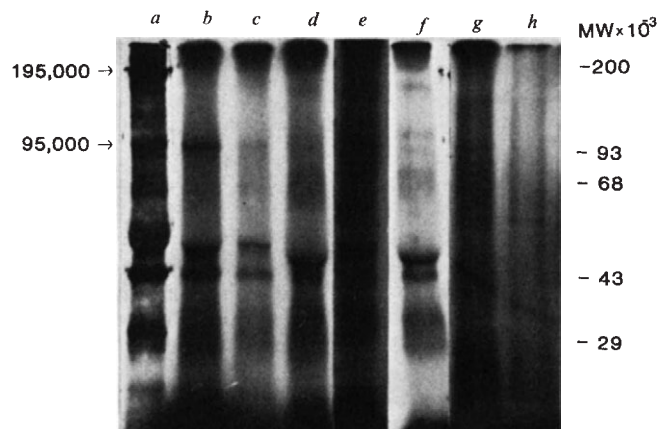


Fig. 3 SDS-PAGE of reduced and alkylated factor B antigen from ^{125}I -labelled mouse peritoneal macrophages radioiodinated for plasma membrane proteins radioiodinated by the Na^{125}I -lactoperoxidase-catalysed reaction¹⁰ with slight modifications. For these experiments, resident and thioglycollate-stimulated peritoneal cells were cultured in serum-free complete DMEM, 60 mM EACA. After culture the cells were released by incubation at room temperature for 1 h with lidocaine (20 mM) in complete DMEM, 20% heated FCS, washed with PBS and then suspended at 1×10^7 cells ml^{-1} for lactoperoxidase radioiodination. The heated FCS contained no detectable factor B antigen. For the cycloheximide experiments, thioglycollate-stimulated cells were cultured for 4 h with cycloheximide added to a concentration of $7.5 \mu\text{g ml}^{-1}$ of complete DMEM. Cycloheximide was removed after 4 h, the cells were rinsed and the culture continued for 36 h in fresh medium containing 60 mM EACA; a control culture contained no cycloheximide. At the end of the culture, the adherent cells were released with lidocaine and radioiodinated for analysis of plasma membrane proteins. The cell suspension was reacted with 250 μCi of carrier-free Na^{125}I , an equal volume of NaI (1 mM solution), lactoperoxidase (40 U; Sigma), glucose oxidase (1 U; Sigma), and the mixture was shaken intermittently at 4°C for 25 min. After incubation, the cells were washed with PBS containing (mM): glucose 20, EDTA 10, PMSF 2, EACA 100 and NaI 75; and finally with complete DMEM, 20% heated FCS containing EDTA, PMSF, EACA and NaI as above. The cells were lysed with cold lysis buffer then centrifuged at 10,000 r.p.m. for 35 min; the supernatants were immunoprecipitated for factor B antigen. Parallel control precipitation was with control antiserum to ovalbumin. The immunoprecipitate was reduced and alkylated and subjected to SDS-PAGE; the autoradiograph is shown. Lane a shows ^{35}S -methionine-labelled factor B antigen from a 48 h culture of thioglycollate-stimulated mouse peritoneal macrophages for comparison with radioiodinated membrane factor B antigen. Note the smaller MW protein bands. Lanes b-e and g, h show lysates of peritoneal macrophages and f shows lysates from non-adherent peritoneal cells. Note (1) the presence of the 95,000-MW protein band in b, representing factor B antigen; (2) the almost total abrogation of the 95,000-MW band in cycloheximide-treated ^{125}I -labelled thioglycollate-stimulated macrophages (c); (3) the absence of the corresponding band from the control precipitate (d); (4) the presence of the corresponding 95,000-MW protein band in the stimulated cells after 36 h of culture in fresh medium following a 4 h exposure to cycloheximide (e); (5) the absence of this band from non-adherent stimulated cells from 48 h cultures (f); (6) the presence of 95,000-MW protein band in radioiodinated resident cells (g); and (7) the absence of the corresponding band in the control precipitate from macrophages (h). ^{14}C -labelled molecular weight markers are shown on the right.

These findings establish the 95,000-MW polypeptide chain detected in cell lysates of mouse macrophages as membrane factor B of molecular weight 95,000, synthesized in short-term primary culture. Secreted factor B of molecular weight 90,000 was also demonstrated. A larger 195,000-MW polypeptide, pro-factor B, has been identified. In studies¹³ using pulse-chase and conversion experiments, the 195,000-MW molecule was shown to be precursor, giving rise to membrane and native factor B. Smaller fragments were also demonstrable and probably result from degradation by proteolytic digestion, as was evident in previous studies⁴. This was probably not due to the specificity of the antibody which was described earlier and is further supported by the presence of only a single band on analysis of immunoprecipitated extracellular material as well as of mouse serum. Another explanation, though less likely, is the presence of incompletely synthesized chains. The biosynthesis of native factor B by mouse macrophages and by human monocytes and the function of plasma factor B have been established previously¹⁴⁻²². The functional activities of the three proteins reported here have yet to be determined. It may be that the macrophage assembles such complement proteins as a defence mechanism, either as a recognition or as an effector unit.

We thank Mrs V Henry for her help. This work was presented at the 9th International Complement Workshop in November 1981 and is supported in part by a grant AI 16729 from the N.I.A.I.D., NIH.

Received 5 January; accepted 5 May 1982.

- Colten, H. R. *Adv. Immun.* **22**, 67-118 (1976).
- Hall, R. E. & Colten, H. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1705-1710 (1977).
- Brade, V., Hall, R. E. & Colten, H. R. *J. exp. Med.* **146**, 759-768 (1977).
- Roos, M. H., Atkinson, J. P. & Shreffler, D. C. *J. Immun.* **121**, 1106-1115 (1978).
- Ooi, Y. M. & Colten, H. R. *J. Immun.* **123**, 2494-2498 (1979).
- Saunders, D. & Edidin, M. *J. Immun.* **112**, 2210-2210 (1974).
- Ferrone, S., Pellegrino, M. A. & Cooper, M. *Science* **193**, 53-55, (1976).
- McConnell, I. & Lachmann, P. J. *Transplant. Rev.* **32**, 72-92 (1976).
- McConnell, I. & Lachmann, P. J. *Immun. Commun.* **6**, 111-132 (1977).
- Yin, H. L., Aley, S., Bianco, C. & Cohn, Z. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2188-2191 (1980).
- Halbwachs, L. & Lachmann, P. J. *Scan. J. Immun.* **5**, 697-704 (1976).
- Woo, P. & Lachmann, P. J. *Immunology* **44**, 629-639 (1981).
- Ooi *et al.* (submitted for publication).
- Bentley, C., Bitterman, D., Hadding, U. & Brade, V. *Eur. J. Immun.* **6**, 393-398 (1976).
- Whaley, K. J. *exp. Med.* **151**, 501-516 (1980).
- Gotze, O. & Muller-Eberhard, H. J. *J. exp. Med.* **134**, 90S-108S (1971).
- Muller-Eberhard, H. J. & Gotze, O. *J. exp. Med.* **135**, 1003-1008 (1972).
- Gotze, O. *Proteases and Biological Control* (eds Reich, G., Rifkin, D. B. & Shaw, E.) 255-265 (Cold Spring Harbor Laboratory, 1975).
- Lesavre, P. H., Hughli, T. E., Esser, A. F. & Muller-Eberhard, H. J. *J. Immun.* **123**, 529-534 (1979).
- Gotze, O., Bianco, C. & Cohn, A. A. *J. exp. Med.* **149**, 372-386 (1979).
- Sundsmo, J. S. & Gotze, O. *Cell. Immun.* **52**, 1-17 (1979).
- Sundsmo, J. S. & Gotze, O. *J. exp. Med.* **154**, 768-777 (1981).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Lasky, R. A. & Mills, A. D. *Eur. J. Biochem.* **56**, 335-341 (1975).

A bifunctional gene product involved in two phases of the yeast cell cycle

J. R. Piggott, R. Rai & B. L. A. Carter

Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland

The cell cycle in *Saccharomyces cerevisiae* is divided into two distinct phases. Unbudded, mononucleate cells in the G1 phase can react to relevant environmental changes by mating¹, sporulating², or by entering stationary phase³. DNA synthesis and bud initiation occur almost simultaneously and mark 'commitment' to the completion of mitosis. Temperature-sensitive mutations at the *cdc28* locus are known to cause arrest in the G1 phase of the cell cycle at the restrictive temperature^{4,5}. Here we show that the *cdc28* gene product is also active in post-G1 cell cycle functions, and that a different property of the gene product may be required for each phase of the cycle in which it acts.

The *S. cerevisiae* strain RJB-1 was obtained during a search for nonsense mutations in the strain FM11 (ref. 6). At 37°C , RJB-1 arrested with the budded, mononucleate morphology

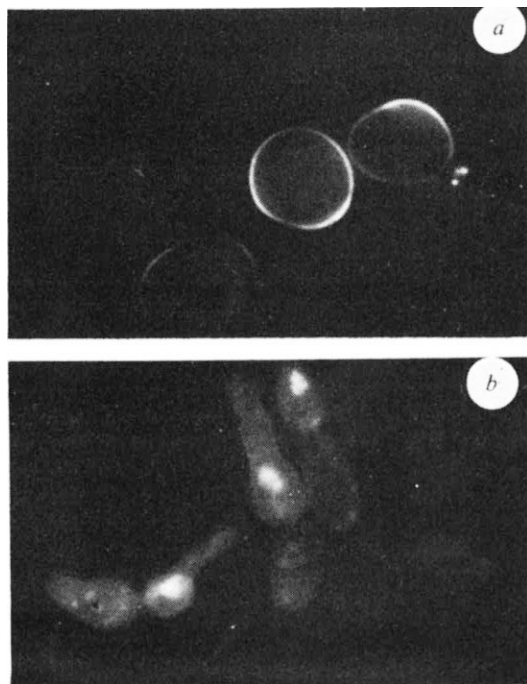


Fig. 1 *a*, The unbudded morphology characteristic of G_1 arrest seen in *cdc28-6* cells after 3.5 h at 37 °C, stained with calcofluor and photographed under UV light. *b*, The mononucleate, budded morphology of RJB-1 (*cdc28-1N*) cells after 3.5 h at 37 °C. The nuclei are stained with diamidinophenyl indole¹⁸ and the cell outlines revealed by calcofluor under UV light.

illustrated in Fig. 1. When crossed with a wild-type strain, RJB-1 yielded a temperature-insensitive diploid, and 42 tetrads from this diploid showed four-spore survival and 2:2 segregation for temperature sensitivity. The homogeneity of its arrest phenotype suggested that RJB-1 carried a cell cycle mutation⁷, and it was shown to complement the *cdc* mutants 1 to 40, except for *cdc28-1* and *cdc28-6*. Dissection of 32 tetrads from the RJB-1 $\times$ *cdc28-1* cross, and 59 tetrads from the RJB-1 $\times$ *cdc28-6* cross, yielded no wild-type spores. We therefore concluded that RJB-1 carried a new mutation at the *cdc28* locus. As the phenotype of the new mutant suggested a lesion in nuclear division, we provisionally named the new *cdc28* allele, *cdc28-1N*.

When yeast cells of *a* mating type are treated with α -factor, a hormone-like small polypeptide produced by cells of α mating type, they undergo reversible G_1 arrest¹. The cells are sensitive to α -factor until ~15 min before the appearance of a visible bud⁸. Figure 2 shows the pattern of DNA synthesis in cells carrying the *cdc28-6* and *cdc28-1N* mutations at 37 °C after release from α -factor-induced arrest at 24 °C. The *cdc28-6*

cells showed no significant DNA synthesis, remaining unbudded, while the *cdc28-1N* cells exhibited appreciable DNA synthesis. Microscopic examination revealed 100% bud initiation in the *cdc28-1N* culture, although there was no increase in cell number. In the reciprocal shift, that is, arrest at 37 °C followed by transfer to 24 °C in the presence of α -factor, *cdc28-6* cells remained arrested in G_1 , as described by Hereford and Hartwell⁹, while *cdc28-1N* cells divided once at 24 °C before undergoing α -factor arrest. Thus, the *cdc28-1N* mutation acts at a point in the cell cycle which is unambiguously later than the α -factor or 'start' step, while the *cdc28-1* and *cdc28-6* execution points cannot be separated from 'start' by the reciprocal shift technique⁹.

The *cdc28-6* $\times$ *cdc28-1N* and *cdc28-1* $\times$ *cdc28-1N* diploids arrested at 37 °C with the budded morphology typical of *cdc28-1N*. As *cdc28-1N* is recessive, this suggested the presence of a nuclear division defect in the G_1 arrest alleles. To test this possibility, α -factor-arrested *cdc28-6* cells were allowed to recover from G_1 arrest at 24 °C, and at 20-min intervals, samples from the recovering culture were shifted to 37 °C. It was found that the cells were unable to divide at 37 °C until 35 min after bud initiation (that is, 'start'). The *cdc28-1N* mutant shows similar behaviour (data not shown). The results presented in Table 1 confirm that *cdc28-6* cells which have recently budded, and therefore passed the 'start' event, cannot complete cell division at the restrictive temperature. Thus *cdc28-6* harbours a lesion in the nuclear division process which has been masked previously by its temporal proximity to the 'start' lesion also present in this strain.

Evidently, mutation at the *cdc28* locus may confer a defect in nuclear division alone, or in both 'start' and nuclear division functions. So far, *cdc28-1N* is the only allele at this locus known to affect nuclear division exclusively. In contrast, a range of *cdc28* G_1 arrest mutants is already available. Whether or not all such mutants show the double cell cycle lesion reported here for *cdc28-6* is unknown. If 'start only' mutants of *cdc28* do exist, we might expect them to complement *cdc28-1N*, and test crosses with this allele should provide a useful preliminary screening procedure. An investigation of the mitotic behaviour of other G_1 arrest mutants (for example, *cdc37*)⁵ is also required in view of these results.

The earliest cell cycle event known to be blocked by the *cdc28* mutation is spindle pole body duplication, which nor-

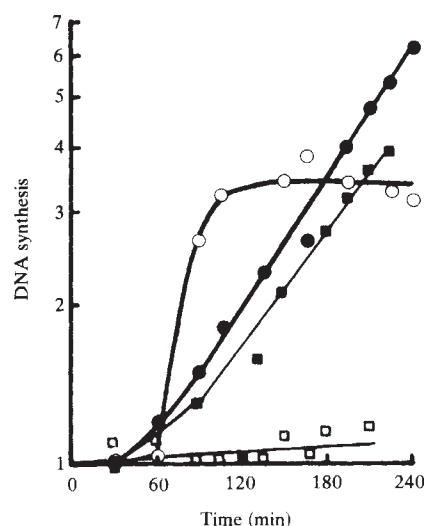


Fig. 2 DNA synthesis patterns following a shift from the α -factor block to 37 °C in RJB-1 (*cdc28-1N*) and ST10 (*cdc28-6*) cells. A log-phase culture (2×10^6 cells ml^{-1}) was arrested in G_1 by exposure for 3 h to α -factor (final concentration 10 U ml^{-1}). The α -factor was removed by Millipore filtration and washing with 100 ml of fresh nutrient medium. Cells were resuspended in fresh medium containing 1 μCi [^3H]thymidine ml^{-1} and incubated at 24 °C or 37 °C. Samples were removed at intervals and the radioactivity incorporated into DNA was determined¹⁹. Results are expressed as c.p.m. relative to initial values in each experiment. ●, RJB-1 at 24 °C; ○, RJB-1 at 37 °C; ■, ST10 at 24 °C; □, ST10 at 37 °C.

Table 1 The behaviour of recently budded *cdc28-6* cells at 37 °C

Time after shift to 37 °C (min)	Unbudded, 1 nucleus (%)	Budded, 1 nucleus in mother (%)	Budded, 1 nucleus in bud (%)	Budded, nucleus dividing (%)	Budded, binucleate (%)
0	48	48	—	—	4
70	33	48	2	4	13
110	34	49	5	—	12
130	32	49	5	—	14

An exponentially growing culture of *cdc28-6* in YEPD (1.5×10^6 cells ml^{-1} in 2% glucose, 2% Bacto-peptone and 1% yeast extract solution) was arrested in G_1 by exposure to α -factor (10 U ml^{-1}) at 24 °C for 3 h. The α -factor was removed by Millipore filtration of the cells, which were then washed with two volumes of fresh medium. The cells were resuspended in fresh YEPD and allowed to recover at 24 °C until 50% had initiated a bud, when a 25-ml sample was shifted to 37 °C. The control cells divided at 24 °C, 100 minutes after T_0 (zero time). Samples taken at the time indicated were stained with diamidinophenyl indole¹⁸ to reveal nuclear morphology. The increase in per cent binucleate cells was due to cells which had already passed both cell cycle blocks at the time of the shift. At least 150 cells were scored for each sample.

mally precedes both initiation of DNA synthesis and budding¹⁰. As recently budded cells have already duplicated their spindle pole bodies, our results indicate that *cdc28* cannot be concerned exclusively with such duplication. Failure to complete this process does not suppress other 'start' activities in the cell cycle mutant *cdc31* (ref. 11), thus it is possible that such failure is a symptom of G₁ arrest, but not its primary cause. As bud formation can occur in the absence of DNA synthesis in *cdc7* (ref. 12), and DNA synthesis can occur in the absence of bud formation¹³, it is likely that all three markers of the start event are normally subject to coordinate control by a process or structure which involves the *cdc28* gene product.

When recovering from α -factor arrest at 24 °C, *cdc28-6* and *cdc28-1N* cells showed nuclear migration into the bud before nuclear division, and this migration was inhibited at 37 °C (Table 1). As mid- and late-nuclear division require extensive nuclear mobility, and these events are apparently unaffected by *cdc28* mutations, it is unlikely that the early nuclear paralysis seen in both *cdc28-6* and *cdc28-1N* is due to a defect in microtubule formation. Nuclear division, however, requires that nuclear movements be spatially coordinated within the mother/bud junction of the dividing cell. Such coordination requires the presence of nuclear/cytoplasmic connections, and the mitotic spindle in yeast is known to connect with the cytoplasm via microtubules¹⁴. Inability to form such nuclear/cytoplasmic connections at 37 °C could well account for our results, and for the reported association of a *kar* phenotype (failure of nuclear fusion at zygote formation) with some *cdc28* mutants¹⁵.

Nuclear division is prevented by the *cdc28-1N* mutation at the restrictive temperature, but start functions seem unaffected. This suggests that a different property of *cdc28* is required for each process in which it participates. The bifunctional nature of the *cdc28* protein may reflect functional segregation of the *cdc28* gene similar to that found at the *his4c* locus, where a single polypeptide has three enzymatic functions which can be independently eliminated by mutation¹⁶. On the other hand, the particular lesion present in *cdc28-1N* may be critical for nuclear division, but not start processes, at the restrictive temperature because of a different molecular environment for the gene product at both cell cycle stages. Extensive genetic and physical analysis will be required to elucidate this.

Whatever the physical basis of the bifunctionality of *cdc28*, it is interesting that the *cdc2* gene in *Schizosaccharomyces pombe* has a regulatory effect at start¹⁷ and acts at two points in the cell cycle equivalent to those described here for *cdc28* in *S. cerevisiae*. Although no mutation has so far been reported in *S. pombe* which blocks one function in the cell cycle exclusively, it is possible that the *cdc28* gene of *S. cerevisiae* and the *cdc2* gene of *S. pombe* have equivalent roles, and that in both organisms a functional component of the nuclear division process has a role in regulation commitment to cell division.

This work was supported by grants from the MRC of Ireland and the National Board of Science and Technology.

Received 19 March; accepted 21 May 1982.

1. Bucking-Throm, E., Duntze, W., Hartwell, L. H. & Manney, T. R. *Expl Cell Res.* **76**, 99–110 (1973).
2. Hirschberg, J. & Simchen, G. *Expl Cell Res.* **105**, 245–252 (1977).
3. Johnston, G. C., Pringle, J. R. & Hartwell, L. H. *Expl Cell Res.* **105**, 79–98 (1977).
4. Hartwell, L. H., Mortimer, R. K., Culotti, J. & Culotti, M. *Genetics* **74**, 267–286 (1973).
5. Reed, S. I. *Genetics* **95**, 561–577 (1980).
6. Rai, R. & Carter, B. L. A. *Molec. gen. Genet.* **181**, 556–558 (1981).
7. Hartwell, L. H., Culotti, J. & Reid, B. *Proc. natn. Acad. Sci. U.S.A.* **66**, 352–359 (1970).
8. Carter, B. L. A. *The Cell Cycle* (ed. John, P. C. L.) (Cambridge University Press, 1981).
9. Hereford, L. M. & Hartwell, L. H. *J. molec. Biol.* **84**, 445–461 (1974).
10. Byers, B. & Goetsch, L. *Cold Spring Harb. Symp. quant. Biol.* **38**, 23–131 (1973).
11. Byers, B. *Molecular Genetics in Yeast. Alfred Benzon Symp.* **16** (Munksgaard, Copenhagen, 1981).
12. Hartwell, L. H. *J. Bact.* **115**, 966–974 (1973).
13. Sloat, B. F. & Pringle, J. R. *Science* **200**, 1171–1173 (1978).
14. Robinow, C. F. & Marak, J. J. *Cell Biol.* **29**, 129–151 (1966).
15. Pringle, J. R. & Hartwell, L. H. *Molecular Biology of the Yeast Saccharomyces* (eds Strathern, J. et al.) (Cold Spring Harbor Laboratory, New York, 1981).
16. Fink, G. R. *Genetics* **53**, 445–459 (1966).
17. Nurse, P. & Bissett, Y. *Nature* **292**, 558–560 (1981).
18. Williamson, D. H. & Fennell, D. J. *Methods in Cell Biology* (ed. Prescott, D. M.) 335–372 (Academic, 1975).
19. Hartwell, L. H. *J. Bact.* **104**, 1280–1285 (1975).

Ligation of oligonucleotides by pyrimidine dimers—a missing 'link' in the origin of life?

Roger J. Lewis & Philip C. Hanawalt

Department of Biological Sciences, Stanford University, Stanford, California 94305-2493, USA

One of the principal photochemical reactions of DNA on exposure to UV is the formation of intrastrand cyclobutane-type pyrimidine dimers^{1,2}. The efficiency of this reaction depends on both the wavelength of the UV² and the specific nucleotide sequence in the DNA³. The formation of the pyrimidine dimer and its repair in living cells have been studied extensively⁴. We have examined the possibility that pyrimidines at the ends of DNA strands may be adequately juxtaposed for dimer formation by the presence of a complementary strand, even when no phosphodiester linkage joins their sugars. In these conditions the formation of a dimer will 'ligate' two DNA strands end-to-end. We report here that thymidine oligonucleotides annealed to polydeoxyadenylate can be ligated end-to-end by UV irradiation, via thymine dimerization of the terminal nucleotides in adjacent oligonucleotides. The linkages are susceptible to direct photoreversal by 254 nm UV, as expected for cyclobutane-type thymine dimers, but they are not cleaved by the bacteriophage T4 endonuclease V, a dimer-specific DNA repair enzyme. We demonstrate that the ligating dimers are also resistant to photolysis from *Escherichia coli*. Although the phosphodiester backbone is not required for dimer formation, it is required for recognition of dimers by these DNA repair enzymes. We discuss the possibility that high molecular weight polynucleotides in primordial seas might have been generated from oligonucleotides by pyrimidine dimerization under the intense solar UV flux unattenuated by an ozone layer.

Within a continuous DNA strand, pyrimidine dimer formation proceeds as follows. An absorbed photon excites a pyrimidine base from its singlet bonding ground state to a singlet antibonding excited state. The new electronic configuration of the base reduces its van der Waals' radius, allowing an adjacent pyrimidine to move closer than the usual 3.4 Å spacing and to share the excitation energy^{5–7}. This 'excimer' state can then undergo inter-system crossing into the more stable triplet state from which a pyrimidine-pyrimidine cyclobutane dimer may arise.

Our model system consisted of oligodeoxythymidylate 10 residues long [d(pT)₁₀] mixed with an equal weight of polydeoxyadenylate [poly(dA)] in solution. The resulting duplex⁸ melts near 25 °C (ref. 9). After labelling a small fraction of the 5' ends with ³²P we irradiated the DNA with UV of wavelengths > 290 nm, which we hereafter call long-wavelength UV. Electrophoretic analysis on denaturing polyacrylamide gels revealed the production of high molecular weight oligomers corresponding to integral multiples of the d(pT)₁₀ (Fig. 1). The presence of dimers, trimers and tetramers was apparent after 30 min of irradiation. This 'ligation' of oligonucleotides was caused presumably by the formation of dimers between the first and last thymines of oligonucleotides adjacent on the complementary poly(dA). Similar irradiation of single-strand d(pT)₁₀ produced no effect detectable by gel electrophoresis (data not shown). We concluded that hybridization to the poly(dA) is required to maintain the oligonucleotides in a photoligatable configuration.

To quantify the amount of material in each electrophoresis band, we excised the bands from the dried gel and assayed them for ³²P. This allowed us to measure the kinetics of the photoligation process (Fig. 2). To compare the photoligation process with the formation of thymine dimers at internal positions, we irradiated poly(dT)·poly(dA) in identical conditions and quantified the resulting thymine dimers by TLC¹⁰. During the initial phase (0–30 min) of irradiation, the production of

dimers, trimers and tetramers of the oligomers occurred, with the expected loss of material from a 10-base band in the gels. The initial rate for internal dimer production in poly(dT)·poly(dA) was similar to or slightly less than the initial rate for the production of ligating thymine dimers in our substrate (Fig. 2).

During the next phase of the process (30–60 min of irradiation) the yield of material in each band reached a plateau. In contrast, the content of internal dimers in poly(dT)·poly(dA) continued to rise rapidly during this period. On continued irradiation (60–240 min) we observed no significant change in the relative quantity of material in the respective bands.

Why does formation of ligating dimers cease while internal dimer formation continues? One possible explanation involves denaturation of oligonucleotides by internal dimers. After 30 min irradiation each oligonucleotide contained an average of one internal dimer, assuming that the rate of internal dimer formation in d(pT)₁₀ is similar to that in poly(dT). A dimer in oligo d(pT)₁₀·poly(dA) reduces the melting temperature below that of d(pT)₆·poly(dA)¹¹. As the melting temperature of d(pT)₆·poly(dA) is less than 9 °C (ref. 9), it seems likely that after 45–60 min of irradiation at 5–10 °C very little of the d(pT)₁₀ was still annealed to the poly(dA), thus eliminating the end-to-end juxtaposition required for the formation of ligating dimers. Another possible explanation is that the competing production of internal dimers between the first and second or ninth and tenth thymines within the oligonucleotides prevented the production of ligating dimers between the first and tenth thymines of adjacent oligonucleotides.

One distinctive characteristic of cyclobutane dimers is their reversibility by irradiation with UV of wavelengths shorter than ~260 nm (ref. 2). To confirm that the ligating dimers were reversible by UV, we exposed d(pT)₁₀·poly(dA) to long-

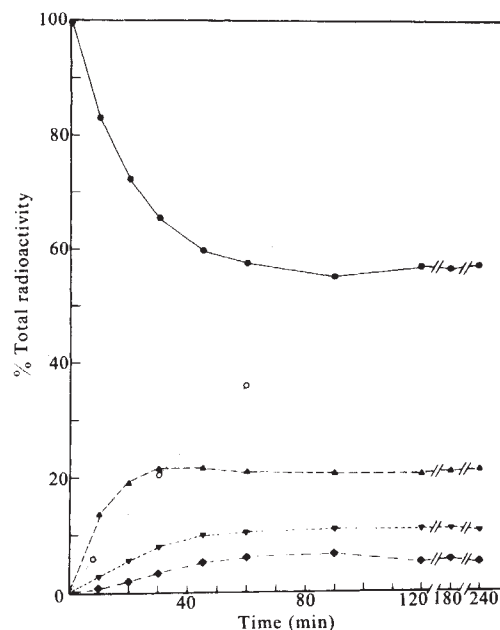


Fig. 2 Kinetics of photoligation. The radioactivity in each band (●, 10 bases; ▲, 20 bases; ▼, 30; ◆, 40) is expressed as a percentage of the total radioactivity entering each gel lane shown in Fig. 1. After the gel was autoradiographed, each band was excised and assayed for radioactivity by scintillation spectrometry in omnifluor/toluene (NEN). A separate experiment was set up to determine the kinetics of internal dimer formation in our irradiation conditions: a small amount of [2-¹⁴C]poly(dT)·poly(dA) (10472; P. L. Biochemicals); was included in the cuvette during an identical irradiation of d(pT)₁₀·poly(dA) and the fraction of thymines contained in dimers was measured by TLC¹⁰. The percentage of total ¹⁴C contained in the dimer peak is shown (○). The initial rate of production for any of the products measured was determined by measuring the slope of the line joining the first two points (for example, 0 min and 10 min).

wavelength UV to produce 'ligated' oligonucleotides, then re-irradiated the DNA with UV of shorter wavelengths (mainly 254 nm). With doses of short-wavelength UV up to 1.6×10^4 J m⁻², the amount of material in the ligated oligomer bands clearly decreased (Fig. 3). Quantification of the material in the bands confirmed this (not shown). This observation demonstrates photoreversibility of the ligating dimers and, as discussed below, lends support to the hypothesis that dimer-induced denaturation limits formation of ligating dimers.

Short-wavelength UV (254 nm) can dimerize up to 39% of the thymines in poly(dT)₂, but only 20% of the thymines were dimerized after our initial long-wavelength UV irradiation of d(pT)₁₀·poly(dA). Thus, the subsequent short-wavelength irradiation while apparently breaking the ligating dimers, should have actually increased the frequency of internal dimers within the oligomers. This dichotomy is understandable if one considers two effects of short-wavelength UV on a photoligated oligonucleotide. First, the internal dimers are both broken and formed, eventually resulting in a higher equilibrium frequency than originally present. This leads to denaturation of the oligonucleotide. Second, the ligating dimer is broken, but cannot be re-formed because the oligonucleotide is no longer hybridized to the poly(dA). Thus there is a loss of ligated oligonucleotides. Dimer-induced denaturation removes the oligonucleotides from a ligatable configuration after either extensive long-wavelength irradiation or after a shorter period of long-wavelength irradiation followed by short-wavelength irradiation. In the former case this simply stops the photoligation process, while in the latter case it allows us to observe photoreversibility of the ligating dimers.

It was of interest to determine whether enzymes specific for pyrimidine dimers in DNA would also recognize the ligating dimers, as this would elucidate both the nature of the ligating dimers and the mechanisms of action of these enzymes.

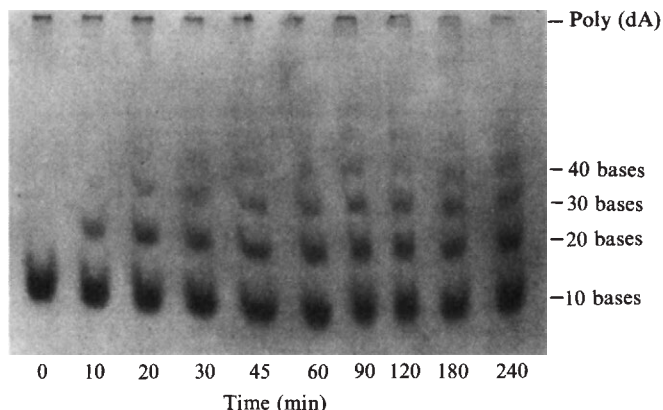


Fig. 1 Ligation of oligonucleotides by increasing irradiation with long-wavelength UV. This autoradiograph of a denaturing polyacrylamide gel shows the production of dimers (20 bases), trimers (30 bases), and tetramers (40 bases) of oligonucleotides (10 bases) by increasing irradiation with long-wavelength UV. Double-strand polymers of oligo d(pT)₁₀·poly(dA) (7850; P. L. Biochemicals) were dissolved in 0.5 ml phosphate buffer (50 mM sodium phosphate, 50 mM NaCl, pH 7.0) to an A₂₆₀ of 6.0; 1 µg of this DNA was ³²P-labelled at the 5' termini²⁰. Mixing the unlabelled and labelled DNA yielded an activity of 10³ c.p.m. µl⁻¹ although less than 1% of the molecules were labelled. For irradiation, a fraction of the DNA solution was placed in a 1 cm glass cuvette, surrounded by water at 0 °C. The solution was stirred and nitrogen was bubbled through it before and during the irradiation. Light was provided by a focused 75 W low-pressure xenon lamp having an output of 2 J m⁻² in the 250–260 nm band. The glass cuvette was opaque to wavelengths shorter than 290 nm and in these conditions a high equilibrium dimer yield was obtained². At the times shown, 30-µl samples were removed, denatured with formaldehyde²¹, and loaded onto the gel with 6 µl of 40% (w/v) sucrose. The 12.5% polyacrylamide gels were prepared and run in Peacock's buffer²² to which formaldehyde had been added to a concentration of 2% (w/w). The buffer was recirculated during both the pre-run (> 8 h, 100 V) and the run (60 min, 250 V). The gel was dried and autoradiographed.

Photolyase from *E. coli* binds tightly to pyrimidine dimers and, on irradiation with visible light, catalyses the splitting of the dimers into the two original pyrimidines^{12,13}. The splitting of a ligating dimer would result in the separation of the oligonucleotides as no phosphodiester bond connects them.

To determine whether the ligating dimers were sensitive to photolyase, we treated photoligated d(pT)₁₀·poly(dA) with the enzyme. We included in the reaction mixture a small amount of ¹⁴C-labelled *E. coli* DNA, the natural substrate, to confirm that the enzyme was active. After electrophoretic analysis we looked for a reduction in the amount of material in the 20-base band which would indicate splitting of the ligating dimers. We monitored splitting of pyrimidine dimers in *E. coli* DNA by TLC. In reaction conditions suggested by Betsy Sutherland (personal communication), 33% of the dimers in the *E. coli* DNA were split but we observed no significant reduction in the relative quantity of 20-base material. Although this suggested that the ligating dimers were insensitive to the enzyme, the possibility remained that the 20-base oligonucleotides were too small for efficient substrate recognition. Competition experiments have shown that photolyase from yeast will efficiently recognize dimers on oligonucleotides 20 base pairs (bp) long¹⁴. To interpret our negative result unambiguously we needed to demonstrate that the *E. coli* enzyme also recognized dimers on small oligomers. In a series of competition experiments analogous to those performed with the yeast photolyase, we demonstrated that irradiated oligo(d(pT)₁₀·poly(dA)), containing photoligated oligonucleotides 20 bp long, did in fact compete with irradiated *E. coli* DNA for the *E. coli* photolyase. Thus, photolyase was able to recognize the dimers in photoligated oligonucleotides 20 bp long and possibly in even smaller oligonucleotides. As photolyase did not break the linkages, we concluded that the lack of a phosphodiester bond at such sites reduces significantly the activity of the enzyme.

Another enzyme that is specific for the cyclobutane pyrimidine dimer is endonuclease V from *E. coli* infected with bacteriophage T4¹⁵. This enzyme (T4 endo V) incises DNA containing pyrimidine dimers by a two-step process. First, a dimer-specific DNA glycosylase cleaves the 5' pyrimidine of the dimer from its sugar and then an apyrimidinic-apurinic endonuclease cuts the phosphodiester linkage between the dimerized bases, leaving an apyrimidinic site at the 3' end at the nick^{3,16}. As there is no phosphodiester linkage between the thymidines in the ligating dimer, the action of the glycosylase alone should break the linkage. We wished to compare the glycosylase activity of T4 endo V on ligating dimers with that on internal dimers. We irradiated d(pT)₁₀·poly(dA) with long-wavelength UV to produce both internal and ligating dimers.

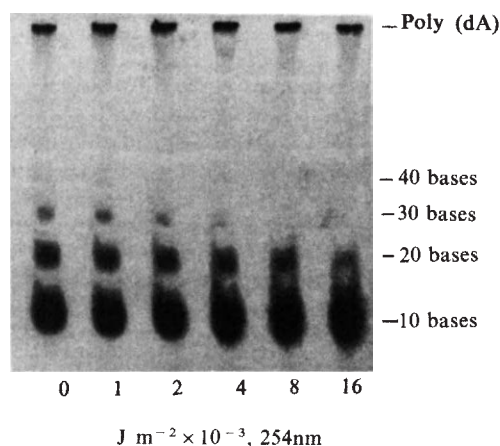


Fig. 3 Photoreversal of ligating dimers by 254 nm irradiation. d(pT)₁₀·poly(dA) was irradiated for 30 min with long-wavelength UV as described in Fig. 1 legend. The solution was then irradiated again at a distance of 10 cm from a 15 W mercury germicidal lamp whose principal output is 254 nm. At times corresponding to the doses shown, samples were removed and analysed by electrophoresis in the same conditions as before.

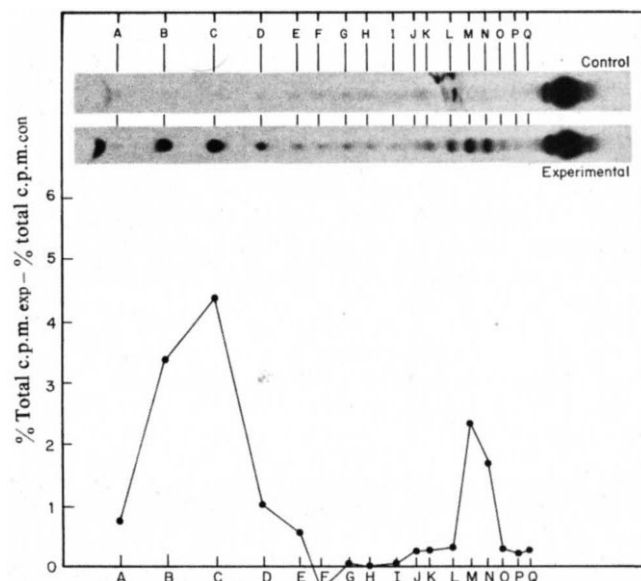


Fig. 4 Action of the glycosylase of T4 endo V on 20-base oligonucleotides containing a ligating dimer between the 10th and 11th positions. The top of the figure shows the electrophoresis patterns of photoligated oligonucleotides before and after treatment with T4 endo V and NaOH. The DNA moved from right to left. The very intense band at the origin represents the 20-base substrate and the bands to the left represent scission products of decreasing size. The top lane shows the pattern obtained for DNA treated with NaOH to hydrolyse the phosphodiester bond at apyrimidinic sites already present. The defect in the autoradiogram at the band labelled L was not the result of radioactivity in the gel. The bottom lane shows the pattern obtained for DNA that had been treated with T4 endo V followed by NaOH. Strand scissions occurred on both ends of the molecule (A-D and L-N). Although oligonucleotide markers were run in adjacent lanes (not shown), unambiguous assignment of lengths to each band was impossible, thus we have only lettered them. Action of the enzyme on the ligating dimer would have yielded a fragment migrating between I and K. Each band shown at the top of the figure was excised and assayed for radioactivity. After expressing the radioactivity of each band as a percentage of the total radioactivity in its lane, we subtracted the control values (c.p.m._{con}) from the experimental values (c.p.m._{exp}). No strand scissions occurred at the centre of the oligonucleotide although every molecule contained a ligating dimer at its centre. The substrate molecules for this experiment were isolated by electrophoresis (12% acrylamide, 7 M urea gel) of irradiated d(pT)₁₀·poly(dA) (20 min long-wavelength UV), followed by electroelution of the 20-base band. We 5'-end-labelled these molecules with ³²P and precipitated them with ethanol. Enzyme reactions were carried out in 20 µl buffer (10 mM Tris, 10 mM EDTA, pH 8.0) at 37 °C for 30 min. NaOH was then added to 0.1 M and the samples were incubated at 90 °C for 30 min. After lyophilization the samples were resuspended in 5 µl formamide and run on a 20% acrylamide, 7 M urea gel²⁰.

After separation on a urea gel, we recovered the 20-base oligomers by electroelution and then ³²P-labelled their 5' ends. Each 20-base oligonucleotide thus obtained contained a ligating thymine dimer between positions 10 and 11 and an average of one or two dimers randomly distributed over the other positions (except positions 9–10 and 11–12, which are prohibited in these molecules). The sample was then treated with excess T4 endo V, and heated in NaOH to ensure that strand scission would occur at all sites of glycosylase activity of the enzyme. A control sample was treated with NaOH after incubation without the enzyme. All enzyme reactions were performed on single-strand DNA to eliminate the possibility that denaturation differences between the locations of internal and ligating dimers might affect the results. The products of the reactions were analysed by electrophoresis on a 20% polyacrylamide DNA sequencing gel (Fig. 4); a band should appear for each dimer site sensitive to the enzyme plus NaOH. As there is a 100% probability of

having a ligating dimer in the 10–11 position but a less than 10% chance of having an internal dimer at any other particular site, this experiment detects susceptibility of the linking dimer to the enzyme with a 10-fold greater sensitivity than that for internal dimers. Nevertheless, no breakage at the position of the linking dimers was detected (Fig. 4). In our laboratory it has been found that thymine dimers in d(pT)₁₇, treated in exactly the same manner, are sensitive to T4 endo V along the entire mid-section of the molecule (to be published), which excludes the possibility that the mid-section of an oligonucleotide is insensitive to the enzyme.

Thus, as was suggested with photolyase, the glycosylase activity of the T4 endo V seems to require a phosphodiester backbone. In this case, however, we demonstrated that internal dimers in our 20-base molecule were sensitive to the enzyme, while the ligating dimer was not. Proposed mechanisms for either enzyme must eventually take into account this feature of substrate recognition and action.

Could the pyrimidine dimer have served as a primordial 'ligase', linking short oligomers of nucleotides for the abiogenic assembly of high molecular weight duplex nucleic acids? In 1974 Tóó pointed out that in the laboratory "no satisfactory chemical synthesis of long-chain polynucleotides has succeeded under any conditions, prebiotic or otherwise"¹⁷. Since then, much progress has been made in template-directed non-enzymatic polymerization of nucleotide derivatives. For example, Zn²⁺ has been shown to serve as "an efficient catalyst for the oligomerization of guanosine 5'-phosphorimidazolide on a polycytidylic acid template"¹⁸.

Once high molecular weight polynucleotide templates existed, simple catalysts could make multiple complementary copies of them. But how did these templates originate and how did they increase in length? Small oligomers might have been created by a number of reactions: for example, condensation of 5' thymidine monophosphates by heating in the presence of histidine has been shown to yield oligomers up to 7 units long and containing predominantly the 3'-5' phosphodiester linkage¹⁹. The yields from such reactions fall off rapidly as the length of the oligomer increases, however, thus it is highly unlikely that the 'first' high molecular weight template formed exclusively by these reactions. Instead, we envisage the accumulation of overlapping oligomeric structures, followed by some ligation event(s) to generate much larger structures. This would occur simultaneously with template-directed polymerization reactions. The enhanced thermal stability of the larger duplex structures should help to drive the process forward.

The results presented here show that the ligation of short oligomers having thymidine nucleotides at their termini could have been accomplished by formation of cyclobutane pyrimidine dimers. In the primordial era there was no stratospheric ozone layer to attenuate the intense UV flux from the sun. Thus, it is fairly certain that pyrimidine-containing oligonucleotides accumulated dimers. We suggest that the cyclobutane pyrimidine dimer that poses a threat to DNA replication and survival in present day organisms, might have played a part in the prebiotic assembly of the first genetic material.

We thank Betsy Sutherland and Pat Seawell for supplying photolyase and T4 endo V, respectively; Ann Ganesan, Michel Philippe, Allen Smith and Mimi Zolan for support; and Evelyn Gates for repeatedly typing the text. This work was supported by a grant GM09901-20 from the National Institute of General Medical Sciences, USPHS.

Received 5 February; accepted 17 May 1982.

1. Smith, K. C. & Hanawalt, P. C. *Molecular Photobiology* (Academic, New York, 1969).
2. Deering, R. A. & Setlow, R. B. *Biochim. biophys. Acta* **68**, 526–531 (1963).
3. Haseltine, W. A. *et al. Nature* **285**, 634–641 (1980).
4. Hanawalt, P. C., Cooper, P. K., Ganesan, A. K. & Smith, C. A. *Rev. Biochem.* **48**, 783–836 (1979).
5. Wang, Shih Yi. *Photochemistry and Photobiology of Nucleic Acids* Vol. 2, 430 (Academic, New York, 1976).
6. Eisinger, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **55**, 1015–1020 (1966).
7. Rahn, R. O. & Landry, L. C. *Photochem. Photobiol.* **18**, 29–38 (1973).
8. Spencer, J. H. *The Physics and Chemistry of DNA and RNA*, 92 (Saunders, Philadelphia, 1972).

9. Janik, B. *Physicochemical Characteristics of Oligonucleotides and Polynucleotides*, 108–109 (Plenum, New York, 1971).
10. Reynolds, R. J., Cook, K. H. & Friedberg, E. C. in *DNA Repair: A Laboratory Manual of Research Procedures* (eds Friedberg, E. C. & Hanawalt, P. C.) 11–22 (Dekker, New York, 1981).
11. Haynes, F. N., Wilkins, D. L., Ratliff, R. L., Varghese, A. J. & Rupert, C. S. *J. Am. chem. Soc.* **93**, 4940–4942 (1971).
12. Sutherland, J. C. *Photochem. Photobiol.* **25**, 435–440 (1977).
13. Snapka, R. M. & Sutherland, B. M. *Biochemistry* **19**, 4201–4208 (1980).
14. Setlow, J. K. & Bollem, F. J. *Biochim. biophys. Acta* **157**, 233–237 (1967).
15. Friedberg, E. C. *Mutat. Res.* **15**, 113–123 (1972).
16. Seawell, P. C., Smith, C. A. & Ganesan, A. K. *J. Virol.* **35**, 790–797 (1980).
17. Tóó, P. O. *Basic Principles in Nucleic Acid Chemistry* Vol. 1 (Academic, New York, 1974).
18. Bridson, P. K. & Orgel, L. E. *J. molec. Biol.* **144**, 567–577 (1980).
19. Ibanez, J. D., Kimball, A. P. & Oro, J. *Science* **173**, 444–446 (1971).
20. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
21. Haselkorn, R. & Doty, P. *J. biol. Chem.* **236**(10), 2738–2745 (1961).
22. Peacock, A. C. & Dingman, C. W. *Biochemistry* **6**(b), 1818–1827 (1967).

Potential Z-DNA forming sequences are highly dispersed in the human genome

Hiroshi Hamada & Takeo Kakunaga

Cell Genetics Section, Laboratory of Molecular Carcinogenesis, National Cancer Institute, NIH, Bethesda, Maryland 20205, USA

Recent analyses of the three-dimensional structure of synthetic DNA molecules has revealed the existence of a left-handed double-helical conformation of DNA, called Z-DNA^{1–4}. The Z-DNA structure was first observed in molecules having alternating guanine and cytosine bases^{1–3}, but has now also been shown for molecules of sequence poly(dT-dG)·poly(dC-dA) (refs 4–7). If Z-DNA occurs naturally then it might have quite different reactivities with molecules such as proteins or carcinogens from right-handed B-DNA. The interconversion of sequences between B and Z forms, under the influence, for example, of DNA binding proteins or chemical modification, may be important in regulating DNA function^{8,9}. So far, little data have been published on the occurrence of potential Z-DNA forming sequences in eukaryotic DNA. Here we report the presence of a sequence of 50 alternating dT and dG residues within one of the introns of a human cardiac muscle actin gene. Also, using a probe specific for poly(dT-dG) sequences, we have found that potential Z-DNA forming sequences are highly repeated in the human genome.

We have recently isolated a phage clone that contains a human cardiac muscle actin gene, λHa-25 (Fig. 1). This clone was obtained by screening of a human DNA library that had been constructed by partial *EcoRI* digestion of DNA from a β⁰-thalassaemia patient and ligation to Charon 4A phage vector¹⁰. The molecular structure (Fig. 1a) and nucleotide sequence of the entire coding region and some introns of the cloned cardiac muscle actin gene were determined¹¹. In intron IV we found a sequence of 25 alternating dT and dG residues (abbreviated to (dT-dG)₂₅). Figure 1b shows the nucleotide sequence of the intron IV–exon 5 boundary region. The presence of (dT-dG)₂₅ near the intron–exon junction was confirmed by sequencing of both DNA strands by the Maxam–Gilbert method. The repeated sequences found are not an artefact of our cloning procedure because we used RecA[–] mutant *Escherichia coli* as a host, and we have further been able to isolate human cardiac muscle actin gene with exactly the same restriction map as that of λHa-25 by screening a second human DNA library constructed by partial *EcoRI* digestion of DNA from a chemically transformed human fibroblast cell line (our unpublished data). Our finding of a (dT-dG)₂₅ stretch is a direct demonstration of potential Z-DNA sequences in the eukaryote DNA. If poly(dT-dG) sequences are functionally important then it is to be expected that they should be found in other parts of the genome. To examine this possibility we carried out the following two Southern blotting analyses. In the first analysis we used ³²P-labelled total DNA from human HeLa cells on a probe to detect repeated sequences in the *EcoRI*–*BamHI* 8-kilobase (kb) fragment of the λHa-25. In this method only

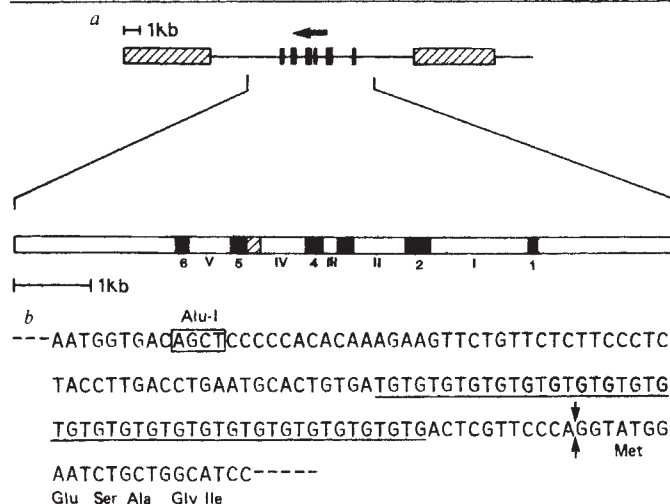


Fig. 1 *a*, Molecular organization of the human cardiac muscle actin gene. The black blocks show the locations of six exons. The arrow above the map indicates the direction of transcription. Six exons and five introns are designated from 1 to 6 and from I to V, respectively, according to their locations. The hatched blocks show the presence of three different repeated elements. These locations were determined by hybridization to ^{32}P -labelled total human DNA or ^{32}P -labelled cloned human *Alu* family (provided by Dr Schmid) (data not shown). *b*, Nucleotide sequences of intron IV-exon 5 boundary region. DNA sequence was determined by the Maxam-Gilbert method²¹. The vertical arrow indicates the splicing site. *AluI* site in intron IV is enclosed by a box. The sequence (dT-dG)₂₅ is underlined. The amino acid sequence corresponding to DNA sequence of exon 5 is shown under the nucleotide sequence.

highly repeated (>100 copies per cell) sequences are detected. When the 8-kb *EcoRI*–*Bam*HI fragment was digested with *Pst*I a 0.9-kb band hybridized to the probe (Fig. 2). A probe specific for the *Alu* family of repetitive sequences¹² did not hybridize to any region of the *EcoRI*–*Bam*HI fragment (data not shown). The 0.9-kb fragment contains all of exon 5 and a part of introns IV and V. In order to narrow further the location of the repeated element, we constructed a subclone of the 0.9-kb *Pst*I fragment in pBR322. By Southern blotting analysis of restriction digested 0.9-kb DNA fragment using ³²P-labelled total human DNA as a probe, the repeated element was mapped between the *Alu*I site in intron IV and *Sau*96I site in exon 5 (Fig. 2). As the actin-coding sequence is not highly repeated in the human genome, we conclude that a highly repeated element is located in the 110-base pair (bp) region between the *Alu*I site in intron IV and the 5' splicing site—precisely the region where (dT–dG)₂₅ is located.

In the second analysis, the DNA used as a probe, and that probed were reversed. ^{32}P -labelled 0.9-kb *Pst*I fragment was used as a hybridization probe against *Eco*RI-digested human DNA. As shown in Fig. 3a a smear of hybridization was observed with the HeLa DNA, indicating again that the 0.9-kb *Pst*I fragment contains the highly repeated element. To demonstrate that the repeated element in the 110-bp region between *Alu*I site and splicing site is the sequence (dT-dG)₂₅, we prepared ^{32}P -labelled poly(dT-dG)·poly(dC-dA) as a probe to specifically recognize sequences. The specificity of the probe was confirmed by the exclusive hybridization of the probe to (dT-dG)₂₅-containing fragments when hybridized to restriction enzyme digested λ HA-25 DNA (Fig. 3b). This probe also produced a smear of hybridization in gels containing *Eco*RI-digested total human DNA (Fig. 3b), thus indicating that (dT-dG)_n is a repeated element. We also determined the copy number of (dT-dG)_n sequences in the human genome by Dot analysis¹³, using the (dT-dG)_n-specific probe (Fig. 4). The approximate number of (dT-dG)_n copies was estimated as 10⁵ per human genome, assuming that the average size of *n* is 25.

These results clearly indicate the widespread distribution in the human genome of sequences that cross-hybridize with the

poly(dG-dT)·poly(dA-dC) probe. The following observations suggest that most of the hybridizing sequences are poly(dT-dG) stretches. First, the hybridization conditions used were so stringent that the sequences containing more than 15% of mismatched nucleotide sequence would not cross-hybridize with the probes. Second, large excess amounts of *E. coli* DNA did not compete at all with the probes for hybridization to human DNA. Third, sequence analysis of one of hybridizing clones containing *Eco*RI-digested human DNA proved to contain an exact (dT-dG)₁₅ sequence (unpublished data). Fourthly, a (dT-dG)₁₇ sequence has been found in human globin genes¹⁴. The observation that our probe did not hybridize to pBR322 DNA (data not shown), which contains a (dT-dG)₃ sequence indicates that the DNA fragments that cross-hybridized to the probe contain a significant stretch of alternating dT and dG residues.

It has been reported that the C(8) position of deoxyguanosine residues, which is the primary position of binding of mutagens and carcinogens such as *N*-acetoxy-*N*-2-acetylaminofluorene (*N*-AcO-AAF), is exposed on the outer surface of the DNA molecules in Z-DNA but not in B-DNA. *N*-AcO-AAF modification or methylation of cytidine in the 5-position in poly(dG-dC)·poly(dG-dC) has been found to favour the transition of the polymer from the B to Z conformation¹⁵⁻¹⁷. Mutational 'hot spots' have been shown to exist in the *hst D* gene of *Salmonella*, within which is a (dG-dC)₄ sequence¹⁸. Further, cloned segments of poly(dG-dC)₁₃·poly(dG-dC)₁₃ in a plasmid undergo a B to Z transformation even though they are enclosed in plasmid segments containing B-DNA¹⁹. Nordheim *et al.* recently reported that antibodies specific to the Z-DNA conformation bind to interband regions of *Drosophila* polytene

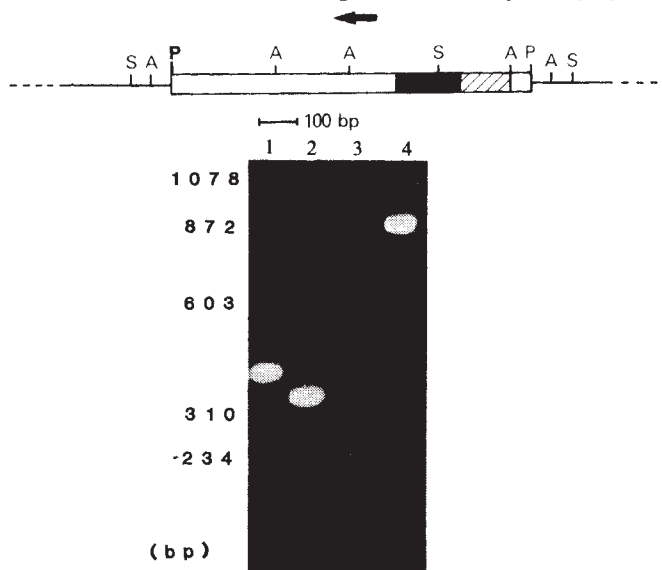


Fig. 2 Mapping of highly repeated sequences in human cardiac muscle actin gene. pEB 8.0 DNA, a subcloned DNA containing the *EcoRI*-*BamHI* 8-kb fragment in pBR322, was digested with *PstI* I (lane 4). pPst 0.9-1 DNA, constructed by subcloning of the *PstI* 0.9-kb fragment shown in Fig. 1 into pBR322, was digested with *AluI* (lane 1), *Sau96I* (lane 2), or *Sau96I* and *PstI* (lane 3). The resulting DNA fragments were separated on a 2% agarose gel, transferred to nitrocellulose filter, and hybridized to total HeLa cell DNA labelled with ^{32}P by nick-translation. Hybridization was carried out as described previously²². *E. coli* DNA was used as carrier DNA. The filter was washed in $1\times\text{SSC } 0.1\%$ SDS at 50°C three times and in $0.1\times\text{SSC } 0.1\%$ SDS at 50°C twice. The restriction map of pPst 0.9-1 DNA is shown above the autoradiogram. The solid and broken lines indicate pBR322 DNA. The box indicates the 0.9-kb *PstI* fragment subcloned into pBR322. The black box and open boxes denote coding region and introns, respectively. The hatched region shows the location of the repeated sequence as determined by the results shown in this figure. The arrow above the map indicates the direction of transcription. The restriction enzyme sites are abbreviated as the following: A, *AluI*; P, *PstI*; S, *Sau96I*. *HaeIII* digests of ΦX174 DNA served as a molecular size standard.

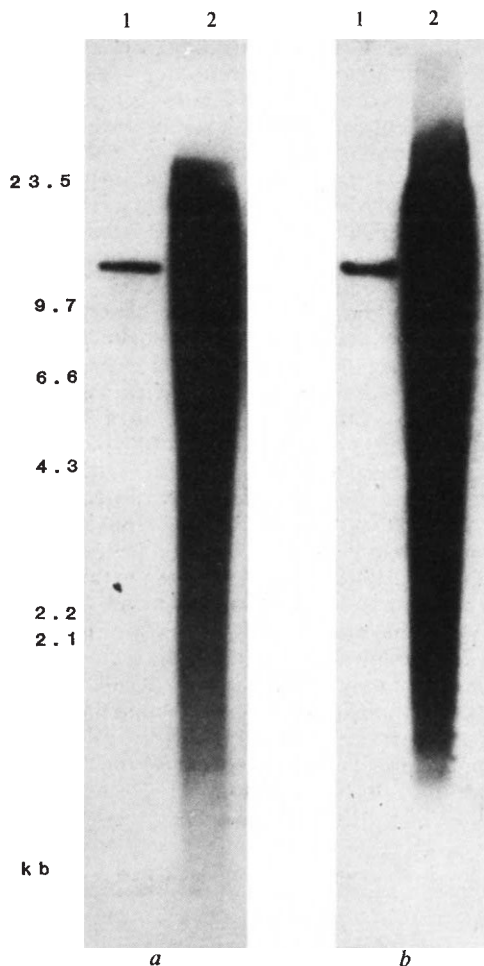


Fig. 3 Detection of poly(dT-dG) sequences in a human genome by Southern blot hybridization. 20 μ g of HeLa cell DNA was completely digested with *Eco*RI (lane 2 in *a* and *b*). Fifty copies equivalent (lane 1 in *a* and *b*) amount of λ Ha-25 DNA was digested with *Eco*RI. These equivalent amounts were calculated based on the assumption that human genome size is 3×10^9 bp, and our data that (dT-dG)₂₅ exist only once in λ Ha-25 (data not shown). The resulting DNA digests were electrophoresed on 0.7% agarose gel, transferred to nitrocellulose filter and hybridized to ³²P-labelled probe. Hybridization was carried out as described above. The probes used here are nick-translated *Pst*I 0.9-kb fragment in *a* and nick-translated poly(dT-dG)·poly(dC-dA) in *b*. The latter probe was prepared by annealing of (dT-dG)₆₋₉ and (dC-dA)₆₋₉ (Collaborative Research) followed by nick-translation.

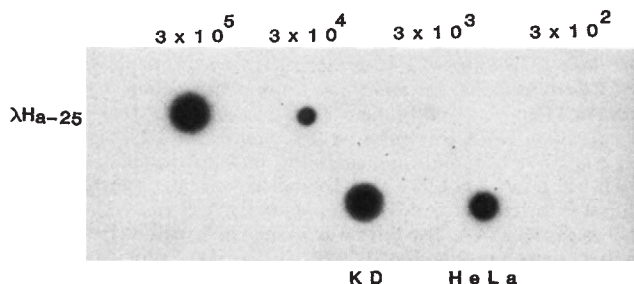


Fig. 4 Determination of copy number of poly(dT-dG) sequences in a human genome. 0.25 μ g of human DNA prepared from HeLa cell or KD cell (normal diploid human fibroblast) and 3×10^5 , 3×10^4 , 3×10^3 or 3×10^2 copies equivalent amount of λ Ha-25 DNA were denatured and spotted on nitrocellulose filter as described by Kafatos *et al.*¹³. These equivalent amounts were calculated as described above. The filter was hybridized to ³²P-labelled poly(dT-dG)·poly(dA-dC) as described above.

chromosomes²⁰. It is plausible that a portion or all of the (dT-dG)_n sequences that we have found to be highly dispersed in the human genome adopt a Z conformation depending on the circumstances and that these sequences may be important in mutagenesis, recombination and the control of gene expression.

We thank Dezider Grunberger and Che-Hung Lee for their valuable suggestions and Kenneth H. Kraemer for his help with and Linda Nischan for typing the manuscript.

Received 29 March; accepted 7 May 1982.

1. Wang, A. H. J. *et al. Nature* **282**, 680-686 (1979).
2. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratliff, R. L. *Nature* **283**, 743-745 (1980).
3. Drew, H., Takano, T., Takano, S., Itakura, K. & Dickerson, R. E. *Nature* **286**, 567-573 (1980).
4. Crawford, J. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 4016-4020 (1980).
5. Leslie, A. G. W., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. *J. molec. Biol.* **143**, 49-72 (1980).
6. Vorlickova, M., Kypr, J., Stokrova, S. & Sponar, J. *Nucleic Acids Res.* **10**, 1071-1080 (1982).
7. Zimmer, C., Tymen, S., Marck, C. & Guschlbauer, W. *Nucleic Acids Res.* **10**, 1081-1091 (1982).
8. Pohl, F. M. & Jovin, T. M. *J. molec. Biol.* **67**, 375-396 (1972).
9. Thamann, T. J., Lord, R. C., Wang, A. H. J. & Rich, A. *Nucleic Acids Res.* **9**, 5443-5457 (1981).
10. Fritsch, E. F., Lawn, R. M. & Maniatis, T. *Cell* **19**, 959-972 (1980).
11. Hamada, H., Petrino, M. G. & Kakunaga, T. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
12. Houck, C. M., Rinehart, F. P. & Schmit, C. W. *J. molec. Biol.* **132**, 289-306 (1979).
13. Kafatos, F. C., Jones, C. W. & Efstradiadis, A. *Nucleic Acids Res.* **7**, 1541-1552 (1979).
14. Misfeld, R., Krystal, M. & Arnheim, N. *Nucleic Acids Res.* **9**, 5931-5947 (1981).
15. Sage, E. & Leng, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4597-4601 (1980).
16. Santella, R. M., Grunberger, D., Weinstein, I. B. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1451-1455 (1981).
17. Behe, M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1619-1623 (1981).
18. Isono, K. & Yournio, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1612-1617 (1974).
19. Klysik, J., Stirdivant, S. M., Larson, J. E., Hart, P. A. & Wells, R. D. *Nature* **290**, 672-677 (1981).
20. Nordheim, A. *et al. Nature* **294**, 417-422 (1981).
21. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
22. Hamada, H., Leavitt, J. & Kakunaga, T. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3634-3638 (1981).

Receptor-rich intracellular membrane vesicles transporting asialotransferrin and insulin in liver

M. T. Debanne, W. H. Evans*, N. Flint & E. Regoeczi

National Institute for Medical Research, Mill Hill, London NW7 1AA, UK, and Department of Pathology, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

A wide range of receptors are located at the blood sinusoidal aspect of the hepatocyte plasma membrane^{1,2}. Many circulating ligands that bind to receptors on the cell surfaces are interiorized along two pathways. Asialoglycoproteins³ are transferred from the plasma membrane to lysosomes and degraded, whereas immunoglobulin A⁴ and bile acids⁵ are transported across the hepatocyte interior and released into bile. Asialotransferrin type 3 (ref. 6) follows a further pathway termed diacytosis⁷. After binding to the asialoglycoprotein receptor⁸, asialotransferrin is endocytosed and then returned to blood with a proportion of its carbohydrate side chains resialylated⁹. We now describe in liver the properties of intracellular asialotransferrin-enclosing vesicles (diacytosomes) and show that they differ from Golgi, lysosome and plasma membrane fractions. Furthermore, we show that the asialoglycoprotein binding sites are located on the cytoplasmic (outer) surface of diacytosomes.

Asialotransferrin, after binding to the hepatic asialoglycoprotein receptor^{3,10-13}, is rapidly interiorized. Analytical centrifugation of liver homogenates in sucrose gradients shows that within 5 s of injecting asialotransferrin into the liver via the portal vein, the ligand equilibrates at a density corresponding to plasma membranes (1.16 g cm^{-3}) and then moves rapidly, through an intermediate density (1.13 g cm^{-3}) at 30 s, to a final density position of $1.10-1.11 \text{ g cm}^{-3}$, where it remains intact for long periods¹⁴. The components equilibrating at the latter

* To whom correspondence should be addressed at the National Institute for Medical Research.

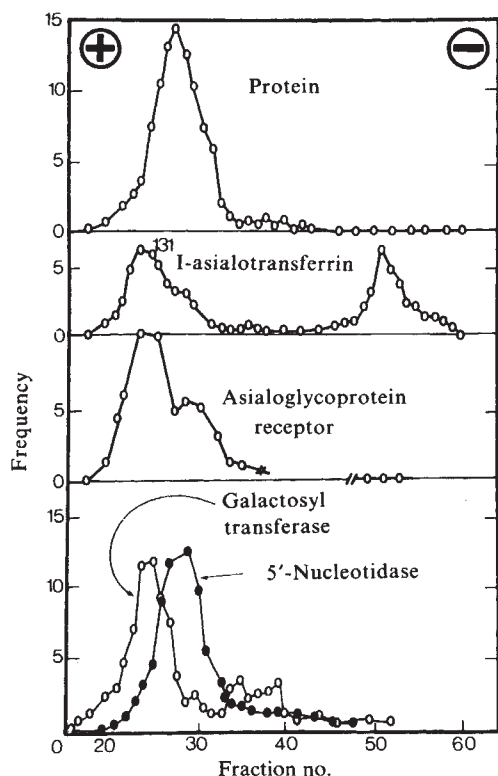


Fig. 1 Fractionation of the diacytosome fraction, prepared as described in Table 1 legend, by free-flow electrophoresis in an Elphor VaP5 apparatus (Bender and Hobein, Munich). The conditions used were essentially as reported by Menashi *et al.*³⁴. Protein was determined by the microtannin turbidimetric method³¹, and asialoorosomucoid binding by the Pricer and Ashwell method³². Galactosyl transferase and 5'-nucleotidase activities were determined by standard methods³⁵. The results are expressed as the frequency of distribution of components recovered from the apparatus for livers 15 min after administration of ¹²⁵I-asialotransferrin.

density were isolated in a subcellular fraction now operationally defined as the diacytosome fraction, using a modified method for preparing blood sinusoidal plasma membranes¹⁵. The diacytosome fraction was obtained (Table 1) by centrifuging the homogenate supernatant in a continuous sucrose gradient after pelleting nuclei, plasma membranes, mitochondria, lysosomes and Golgi apparatus components. Soluble proteins associated with the diacytosome fraction were removed by permeation chromatography. The diacytosome fraction retained about 20% of the bound ¹²⁵I-labelled asialotransferrin and 10–20% of homogenate glycosyl transferase activities (Table 1); recovery of 5'-nucleotidase and alkaline phosphodiesterase was lower and more variable, and that of lysosomal enzymes very low.

On further resolution of the diacytosome fraction by free-flow electrophoresis¹⁶ (Fig. 1), although a single protein peak emerged, the galactosyl transferase and residual 5'-nucleotidase activities separated into adjacent peaks. Importantly, however, the positions of the major asialotransferrin peak, the major peak representing asialoglycoprotein receptor-binding activity and the galactosyl transferase coincided. Although this result could have implied that asialotransferrin was interiorized to the Golgi apparatus, the morphological features and the ligand-receptor binding properties of diacytosomes showed that the recovered ligand resides in previously undocumented subcellular components.

Examination of the diacytosome fraction by electron microscopy showed a relatively homogeneous population of smooth-surfaced vesicular profiles, 600–1,200 Å in diameter. The cisternal tubular networks and vacuoles characteristic of Golgi apparatus fractions^{17–19} were rarely seen. No coated vesicles²⁰ were detected, even when negatively stained preparations were examined. Fractions containing the ligands, receptors and galactosyl transferase (Fig. 1) contained a higher proportion of vesicles with single or compound intraluminal vesicular mem-

Table 1 Recovery and relative specific activities of radioiodinated ligands and marker enzymes in the diacytosome fraction

Component in diacytosome fraction	% Recovery from homogenate	Specific activity relative to homogenate
Protein	0.25 ± 0.04(4)	—
Asialotransferrin type 3	19.4 ± 0.9 (4)	62.0
Asialoorosomucoid	2.7 (2)	6.5
Insulin	7.9 ± 1.1 (4)	25.0
Galactosyl transferase	21.0 (2)	65.0
Sialyl transferase	9.5 (2)	51.0
5'-Nucleotidase	6.7 ± 2.4 (4)	16.5
Alkaline phosphodiesterase	2.5 ± 0.4 (4)	6.5
Acid phosphatase	1.0 (2)	2.2
β-galactosidase	0.09 (2)	0.1

Diacytosomes were prepared from unstarved rat liver homogenates as described in detail elsewhere²². Briefly, nuclei, mitochondria and lysosomes, plasma membranes and Golgi components of livers homogenized in 0.25 M sucrose were removed by centrifugation sequentially at 1×10^4 g min and 4×10^5 g min. The final clear supernatant was centrifuged into a 10–40% w/v sucrose gradient at 97,000g_{av} for 4 h (Beckman SW27 rotor) and components equilibrating at densities of 1.097–1.117 g cm⁻³ were pooled and concentrated by vacuum dialysis. Radioactivity and enzymes (70–90%) and a protein peak (20–25%) were eluted in the void volume peak of a Sepharose 2B column equilibrated in 10 mM Tris-HCl pH 7.4, 1 mM EDTA and 0.15 M NaCl, and concentrated by centrifugation. The three ligands were iodinated using the iodogen method³⁶. Asialotransferrin type 3 (5 µg), asialoorosomucoid (20 µg) or insulin (20 µg) were injected intraperitoneally into rats (150 g body weight) and livers removed and homogenized 10–15 min later when most of the asialotransferrin and proportions of asialoorosomucoid and insulin equilibrated in the density range 1.095–1.12 g cm⁻³. Enzyme activities and radioactivity were determined as previously described^{14,15,24}. The complete distribution of all components in the fractions will be described elsewhere²², and total recoveries were 89–103% of homogenate values. Standard errors of the mean and the number of experiments (in parentheses) are given.

brane components (Fig. 2). Similar vesicles were present as minor components in gross Golgi^{17–19} and plasma membrane sinusoidal fractions¹⁵ prepared from liver. Recently, similar multi-vesicular structures (termed 'receptosomes', although no receptors or receptor binding was measured) were described in cells that have interiorized ligands²¹.

Table 2 shows that the diacytosome fraction has a much higher capacity for binding asialoglycoproteins than do the plasma membrane and Golgi subfractions. Examination of the orientation of the receptors on the vesicles by measuring the binding capacity of various fractions after treatment with Triton X-100 showed that the binding capacity of the Golgi subfractions increased two- to three-fold, indicating that most of the receptor sites were located inside the membrane-enclosed

Fig. 2 Electron micrograph of fraction 23 from the free-flow electrophoresis separation (Fig. 1). Membranes were concentrated by centrifugation on to a 51% (w/v) sucrose cushion, and sections prepared for viewing by standard methods¹⁵. $\times 27,800$.

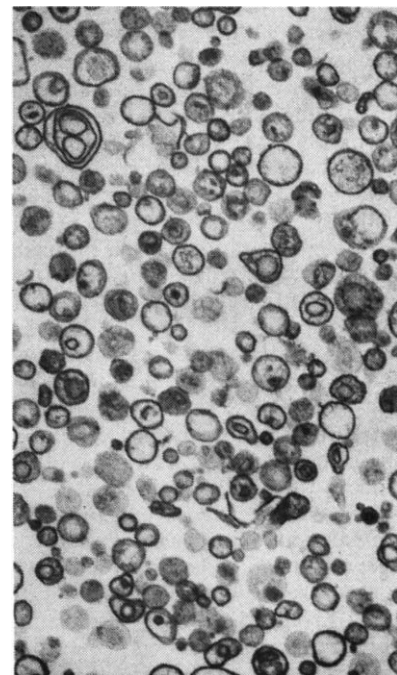


Table 2 Effects of treatment with Triton X-100, trypsin and sonication on binding of asialoorosomucoid and insulin to diacytosome, Golgi and plasma membrane fractions

Fraction	Specific binding activity						Insulin (ng per mg protein)
	Asialoorosomucoid (ng per µg protein)						
	No treatment	0.1% Triton X-100	Sonication	Trypsin	Trypsin and sonication	Trypsin and 0.1% Triton X-100	
Diacytosome	6.1 ± 0.4	5.6 ± 0.6	6.2 ± 0.3	0.03 ± 0.01	0	0.05 ± 0.01	1.14 ± 0.11
Golgi—heavy	0.9 ± 0.2	2.7 ± 0.2	3.1 ± 0.1	0.17 ± 0.1	2.3 ± 0.6	3.3 ± 0.3	0.27 ± 0.02
Golgi—intermediate	0.8 ± 0.1	2.5 ± 0.6	—	—	—	—	0.57 ± 0.09
Plasma membrane (heavy)	1.4 ± 0.2	1.7 ± 0.3	—	—	—	—	0.71 ± 0.02
Plasma membrane (sinusoidal)	2.2 ± 0.3	2.7 ± 0.1	1.6	0.10	0.24	0.03	0.39 ± 0.02

The diacytosome fraction was prepared as described in Table 1 legend, and Golgi and plasma membrane fractions by standard procedures^{15,18}. Asialoorosomucoid binding was carried out as described by Pricer and Ashwell³², and insulin binding as described previously³³. Sonication of membrane fractions was carried out in a Sonic Dismembrator (Fisher, Model 300) using 2 × 10 s bursts at setting '35'. Trypsin treatment was carried out before binding studies as follows. Membrane fractions (250 µg protein) were incubated for 15 min at 23 °C with trypsin (250 µg; P-L Biochemicals); control samples were incubated in the buffer 0.15 M NaCl, 0.005 M CaCl₂, 0.01 M Tris-HCl pH 8. Protease treatment was stopped by addition of a fivefold (w/w) excess of crystalline soybean trypsin inhibitor (Worthington). In separate studies, the trypsin inhibitor had no effect on ligand binding. Values are shown ± s.e.m. of three to five determinations.

compartments. In contrast, no increase in binding capacity was elicited by detergent treatment of the diacytosome fraction, thus indicating that the binding sites were located on the outer surface. Further evidence for an external location of the asialoglycoprotein receptor on diacytosomes was obtained by the finding that sonication had little effect on asialoorosomucoid binding by the diacytosome fraction, whereas a three-fold increase, similar to that elicited by treatment with Triton X-100, was obtained with a Golgi fraction. Trypsin treatment removed 96% of the asialoorosomucoid binding capacity of diacytosomes when carried out alone or in combination with sonication or exposure to Triton X-100, two treatments that should reveal any additional receptors located inside the vesicles. On the other hand, the diminution in binding capacity of the Golgi fraction occurring after trypsin treatment was low when compared with that measured after sonication or Triton X-100 treatment. These results provide strong evidence that the receptor-binding sites are located mainly on the outer (cytoplasmic) surface of the diacytosome.

A role for diacytosomes in the receptor-mediated uptake of insulin was also briefly investigated, for about 8% of the ¹²⁵I-labelled insulin associated with particulate components of the liver homogenate 10–15 min after injection was recovered in the diacytosome fraction (Table 1). Comparison of the insulin-binding capacity of the diacytosome with plasma membrane and Golgi fractions showed that diacytosomes possessed the highest activity (Table 2), and preliminary observations suggest that a proportion of the insulin receptors is cytoplasmically orientated.

The morphology, high concentration and external orientation of receptors, when taken with the simplified polypeptide composition relative to plasma membrane and Golgi fractions²², indicate that diacytosomes cannot be categorized directly into the major hepatic classes of subcellular organelles and membrane systems described^{23,24}. However, in view of the structural and functional complexity of the Golgi apparatus, especially in secretory tissues²⁵ the vesicles may originate from a region of the Golgi apparatus specialized for the processing of classes of ligands internalized by hepatocytes. A diacytotic pathway along which ligands move in and out of the hepatocyte can be divided into inward (endocytotic) and outward (exocytotic) sector legs. The association of glycosyl transferases with the diacytosome fraction indicates that the more electronegative vesicles separated by free-flow electrophoresis may arise from components comprising the intracellular terminus of the diacytotic pathway. The glycosyl transferases in the diacytosome fraction can also explain the glycosylation of internalized asialotransferrin demonstrated in *in vivo* experiments⁹.

Finally, the cytoplasmic orientation of asialoglycoprotein receptors demonstrated on the diacytosomes indicates that the receptor–ligand complex formed at the plasma membrane has been rearranged during internalization. Either there is receptor reorganization in the membrane lipid bilayer²⁶, or engulfment of membrane components of the inward sector leg and multiple

fusions have taken place, resulting in a dissociation of the complex with the receptors now becoming cytoplasmically orientated. Indeed, these results add to the increasing indirect evidence for discrete intracellular compartments bearing cytoplasmic receptors for lectins²⁷, phosphomannosyl enzyme²⁸, asialoglycoproteins²⁹ and human growth hormone³⁰. External receptors on diacytosomes, while undoubtedly contributing to their high electronegativity exploited in the free-flow electrophoretic separation, may provide recognition potential ensuring specificity in the vesicular transport of ligands inside the cell. The high concentration of receptors on diacytosomes suggests a role in receptor recycling and therefore receptor turnover at the plasma membrane.

We thank Professor Neville Crawford and Dr Donald Stuart, Institute of Basic Medical Sciences, Royal College of Surgeons, London for their help and support in running the free-flow electrophoresis apparatus, and Miss Kate Sullivan for electron microscopy. Part of this study was supported by the MRC of Canada and the Ontario Heart Foundation.

Received 11 March; accepted 14 May 1982.

- Evans, W. H. *Biochim. biophys. Acta* **604**, 27–64 (1980).
- Evans, W. H. *Hepatology* **1**, 452–458 (1981).
- Neufeld, E. F. & Ashwell, G. in *The Biochemistry of Glycoproteins and Proteoglycans* (ed. Lennarz, W. H.) 241–266 (Plenum, New York, 1980).
- Renston, R. H., Jones, A. L., Christiansen, W. D., Hradek, G. T. & Underdown, B. J. *Science* **208**, 1276–1278 (1980).
- Jones, A. L., Smuckler, D. L., Mooney, J. S., Ockner, R. K. & Alder, R. D. *Lab. Invest.* **40**, 512–517 (1979).
- Regoezi, E., Taylor, P., Debanne, M. T., März, L. & Hatton, M. W. C. *Biochem. J.* **184**, 399–407 (1979).
- Tolleshaug, H., Chindemi, P. A. & Regoezi, E. *J. biol. Chem.* **256**, 6526–6528 (1981).
- Hudgin, R. L., Pricer, W. J., Ashwell, G., Stockert, R. I. & Morrell, A. G. *J. biol. Chem.* **249**, 5536–5543 (1974).
- Regoezi, E., Chindemi, P. A., Debanne, M. T. & Charlwood, P. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2226–2230 (1982).
- Regoezi, E., Taylor, P., Hatton, M. W. C., Wong, K.-L. & Koj, A. *Biochem. J.* **174**, 171–178 (1978).
- Hubbard, A. L., Wilson, G., Ashwell, G. & Stukenbrok, H. *J. Cell Biol.* **83**, 47–64 (1979).
- Wall, D. A., Wilson, G. & Hubbard, A. L. *Cell* **21**, 79–93 (1980).
- Steer, C. J. & Clarenberg, R. *J. biol. Chem.* **254**, 4457–4461 (1979).
- Debanne, M. T. & Regoezi, E. *J. biol. Chem.* **256**, 11266–11272 (1981).
- Wisher, M. H. & Evans, W. H. *Biochem. J.* **146**, 375–388 (1975).
- Hannig, K. & Heidrich, H. G. *Meth. Enzym.* **31**, 746–761 (1974).
- Ehrenreich, J. H., Bergeron, J. J. M., Siekevitz, P. & Palade, G. M. *J. Cell Biol.* **59**, 45–72 (1973).
- Bergeron, J. J. M. *Biochim. biophys. Acta* **555**, 493–503 (1979).
- Borgese, N. & Meldolesi, J. *J. Cell Biol.* **90**, 687–696 (1981).
- Pearse, B. M. F. & Bretscher, M. A. *rev. Biochem.* **50**, 85–101 (1981).
- Pastan, I. & Willingham, M. C. *Science* **214**, 504–509 (1981).
- Evans, W. H., Debanne, M. T., Flint, N. A. & Regoezi, E. *J. biol. Chem.* (submitted).
- De Duve, C. *Science* **189**, 186–194 (1975).
- Evans, W. H. *Preparation and Characterisation of Mammalian Plasma Membranes* (North-Holland, Amsterdam, 1978).
- Rothman, J. C. *Science* **213**, 1212–1219 (1981).
- Blumenthal, R., Klausner, R. D. & Weinstein, J. N. *Nature* **288**, 333–338 (1980).
- Pinto da Silva, P., Torrisi, M. R. & Kachar, B. *J. Cell Biol.* **91**, 361–372 (1981).
- Fischer, H. D., Gonzalez-Noriega, A., Sly, W. S. & Morré, D. J. *J. biol. Chem.* **255**, 9608–9615 (1980).
- Tanabe, T., Pricer, W. E. & Ashwell, G. *J. biol. Chem.* **254**, 1038–1043 (1979).
- Bergeron, J. J. M., Rachubinski, R. A., Sikstrom, R. A., Posner, B. I. & Paiement, J. J. *Cell Biol.* **92**, 139–146 (1982).
- Mejbaum-Katzenellenbogen, W. & Dobryszczy, W. M. *Clin. chim. Acta* **4**, 515–522 (1959).
- Pricer, W. E. & Ashwell, G. *J. biol. Chem.* **251**, 7539–7544 (1976).
- Evans, W. H., Bergeron, J. J. M. & Geshwind, I. I. *FEBS Lett.* **34**, 259–262 (1973).
- Menashi, S., Weintraub, H. & Crawford, N. J. *J. biol. Chem.* **256**, 4095–4101 (1981).
- Evans, W. H., Flint, N. A. & Vischer, P. *Biochem. J.* **192**, 903–910 (1980).
- Fraker, P. J. & Speck, J. C. *Biochem. Biophys. Res. Commun.* **80**, 849–857 (1978).

MATTERS ARISING

Dielectric constant of ice

ADAMS has recently considered the dielectric constant of ice¹. His Monte Carlo calculations fill a gap in the literature and complement previous calculations. However, his theoretical discussion, while displaying an important constraint on the Kirkwood–Frohlich (K–F) theory, falls short of the ‘simple resolution’ claimed.

Adams’ primary equation¹ is

$$\langle \mu_1 \Sigma \mu_i \rangle_{\epsilon} N / 9 \epsilon_0 k T V \\ = (\epsilon - 1)(2\epsilon' + 1) / 3(\epsilon + 2\epsilon') \quad (1)$$

The right hand side (RHS) of equation (1) is obtained using macroscopic electrostatic theory for a spherical system of ice of dielectric constant ϵ surrounded by a much larger sphere of dielectric constant ϵ' . The left hand side (LHS) of equation (1) is obtained using microscopic statistical mechanics for a system of water dipoles in the central sphere which interact with each other as well as with the material in the outer sphere, whence the subscript ϵ' . The internal field E_{int} which acts on the dipoles in the calculation of the LHS of equation (1) is interpreted to be the cavity field which would have existed if the ice sphere were removed, whereas the polarization P is computed from the field actually present in the electrostatic calculation before the ice sphere is removed. The first problem is to determine specifically which intermolecular interactions are required in the calculation of the LHS to make it correspond to the RHS for which the cavity field is used. This has not been *a priori* obvious as evidenced by the fact that several researchers^{1–7}, including Adams, have performed calculations assuming only short range interactions which Adams correctly concludes is an inconsistent model in the context of equation (1).

For the ideal ice-rules model the numerical results for the correlation sums g_K and G for the LHS appear to agree with the RHS for the respective cases of $\epsilon' = \epsilon$ and $\epsilon' = \infty$. However, as Adams acknowledges, for a finite concentration of Bjerrum defects, which is a necessary feature of real ice, equation (1) becomes inconsistent because statistical mechanics requires that the two correlations sums G and g_K are equal when the ideal ice rules are broken and only short range interactions are included. Adams tries to eliminate this dilemma by the proposal that dipolar interactions, previously left out of all but one primitive calculation⁸, will restore the numerical values of G and g_K to those given by the ideal ice-rules. As Adams admits, this proposal cannot be exact because the ice rules determine

G/g_K to be independent of the dipole moment and temperature which cannot be true for dipolar interactions. He also admits that dipolar interactions are likely to raise the numerical value of G which would reduce the calculated dipole moment below the value $\mu = 3.0$ D quoted in his paper. Therefore, Adams’ simple resolution leaves us with a theory for which neither G nor g_K has been calculated and for which such calculations may be practically impossible due to the difficulty of doing statistical mechanical calculations on systems with long range dipolar interactions. Adams’ real conclusion, also emphasized by Stillinger¹⁰, is that G and g_K must satisfy a relation (Adams’ equation (9)) if the proper interactions consistent with the internal cavity field are included in the calculation of the LHS of equation (1).

Adams’ theory is essentially an extension of the K–F theory to include the cases $\epsilon' \neq \epsilon$ and this allows him to connect the K–F equation, given by equation (1) when $\epsilon' = \epsilon$ for which the LHS has a factor g_K , to an equation which is derived from equation (1) by setting $\epsilon' = \infty$ for which the LHS has a factor which Adams identifies as G . If this last identification is made, then the Onsager–Slater (O–S) equation is recovered and must be equivalent to the K–F equation because they both come from the same unified derivation. However, this identification is incorrect because to calculate the G required by Adams’ theory it is necessary to include dipolar interactions; in contrast dipolar interactions are completely absent in the calculation of G appropriate for the O–S theory. This astonishing feature has caused many to dismiss the O–S theory too lightly. However, elsewhere I have shown that the O–S theory consists of a renormalization of the electrostatic interactions in ice into a dipole-free effective Hamiltonian⁹. Therefore, the G required for the O–S theory cannot be the same as the quantity required for the LHS of equation (1) which should henceforth be called by a different name, $g_{\infty = \epsilon'}$. It then follows that the use of equation (1) does not permit a derivation of the O–S theory which is fundamentally different from the K–F theory^{2,3,9}; in particular, the internal field in the O–S theory is not the cavity field. A major advantage of the O–S theory to the K–F theory is that extant calculations of G , which do not include dipolar interactions, can be applied immediately and unambiguously to the O–S theory. However, it is to be hoped that the more difficult calculation of g_{∞} necessary for the K–F theory may also be performed and that this might indeed result in complete resolution of the theory of the dielectric constant of ice.

J. F. NAGLE

Department of Physics,
Carnegie–Mellon University,
Pittsburgh,
Pennsylvania 15213, USA

1. Adams, D. J. *Nature* **293**, 447–449 (1981).
2. Hollins, G. T. *Proc. phys. Soc.* **84**, 1001–1016 (1964).
3. Nagle, J. F. *J. Glaciol.* **21**, 73–81 (1978).
4. Rahman, A. & Stillinger, F. H. *J. chem. Phys.* **57**, 4009–4017 (1972).
5. Gobush, W. & Hoeve, C. A. J. *J. chem. Phys.* **57**, 3416–3421 (1972).
6. Stillinger, F. H. & Cotter, M. A. *J. chem. Phys.* **58**, 2532–2541 (1973).
7. Yanagawa, A. & Nagle, J. F. *Chem. Phys.* **43**, 329–339 (1979).
8. Powles, J. G. *J. chem. Phys.* **20**, 1302–1309 (1952).
9. Nagle, J. F. *Chem. Phys.* **43**, 317–328 (1979).
10. Stillinger, F. H. in *Studies in Statistical Mechanics* (eds Lebowitz, J. L. & Montroll, E. W.) (North-Holland, Amsterdam, in the press).

ADAMS REPLIES—The ‘simple resolution’ which I claim is the observation that it is not necessary to choose between either the Onsager–Slater equation or the Kirkwood equation. The Onsager–Slater equation, $\epsilon - 1 = 3yG$, becomes equivalent to the exact Kirkwood equation when G is identified with g_{∞} which is obtained with $\epsilon' = \infty$ (that is, the ‘tin-foil’ case) and when all the dipole–dipole and other interactions of the model hamiltonian are included in the calculation^{1,2}. The approximation of the Onsager–Slater theory is that the dipole–dipole interactions are to be ignored in the calculation of G .

Perfect ice-rules ice is a very simple hamiltonian which despite having no dipole–dipole interactions produces long-range dipole–dipole correlations of the sort to be expected in real dielectrics. I have calculated unambiguous values of $G = g_{\infty}$ and g_K for this model. The identification of such a model with real ice must be done with caution and I have tried to make an informed guess as to the direction of the error. One major approximation of the model is that all configurations allowed under the ice rules have the same energy. This cannot be too drastic an approximation or the observed residual entropy of ice would not be predicted so well.³

The other major approximation I have made is to ignore the Bjerrum defects which undoubtedly occur in real ice. However, when non-interacting defects are introduced into the model then the long-range correlations are destroyed and $g_K = G = g_{\infty}$ which cannot be true for a real dielectric. I argue that to introduce defects without also including the long-range electrostatic interactions which occur in all real dielectrics is to make the model worse and not better. I agree with Nagle that complete resolution requires making difficult calculations with all electrostatic interactions included and using the formally exact expressions for the dielectric constant of the Kirkwood type.

D. J. ADAMS

Department of Chemistry,
The University,
Southampton SO9 5NK, UK

1. De Leeuw, S. W., Perram, J. W. & Smith, E. R. *Proc. R. Soc. A* **373**, 27–56 (1980).
2. Pollock, E. L. & Alder, B. J. *Physica* **102A**, 1–21 (1980).
3. Nagle, J. F. *J. math. Phys.* **7**, 1484–1491 (1966).

Body temperature changes during the practice of g Tum-mo yoga

IN their recent letter¹, Benson *et al.* do not provide sufficient information to determine whether the temperature changes reported indeed resulted from the practice of yoga. The 'control' time shown on their graphs ranged from 5 to 10 min, and significant fluctuations in temperature appeared to occur within even that brief time. For example, if the 1°C change in finger temperature recorded during this time for the first subject, V.G.J., were extrapolated for the 55-min meditation period, a temperature change of ~5°C would result, which represents most of the reported effect of 5–9°C. The change in the toe temperature was even greater than the finger temperature change during the 'control' period.

I would like to see results of tests on non-meditating individuals, and on those individuals studied, to determine how their temperatures vary over periods of 1–1.5 hours, corresponding to the duration of meditation. I would also want experiments to be performed in a temperature controlled environment, as I suggest that room temperature changes of 2°C, such as occurred during the trial of the second subject, could lead to greater changes in the temperature of the extremities.

JAY M. PASACHOFF

Williams College,
Hopkins Observatory, Williamstown,
Massachusetts 01267, USA

1. Benson, H. *et al. Nature* **295**, 234–236 (1982).

BENSON *ET AL.*¹ present an interesting account of three persons who increase their finger and toe temperatures apparently by a Tibetan Buddhist meditative practice known as g Tum-mo (heat) yoga. To ensure that any observed changes in body temperature are directly attributable to this method of meditation, the use of a control group is desirable. As this study did not include a control group, it is possible that the observed effects were the result of some other factor, for example: (1) changes in the ambient air temperature (as was the case for two of the subjects, A and B, in this investigation); (2) the consumption of certain food known to increase metabolic activity; and (3) the placement of toe and fingers during the meditation and recovery period, for example, if toes or fingers were placed in an axillary cavity, this would effect a rapid rise in temperature. In addition, the

location and method of attachment of the disk thermistors may be partly responsible for the changes in temperature. Further investigation should include the use of matched controls; possibly additional practitioners of g Tum-mo yoga could be used if they were willing to refrain from meditating during the testing period.

MARC A. STRASSBURG

County of Los Angeles
Department of Health Services,
313 North Figueroa Street,
Los Angeles, California 90012, USA

1. Benson, H. *et al. Nature* **295**, 234–236 (1982).

BENSON REPLIES—With regard to the comments of Strassburg and Pasachoff, the subjects in our experiment served as their own controls in an 'off-on-off' design. As the studies were performed in the isolated, mountain hermitages of the monks, it was not feasible to control the ambient temperature. The monks could not serve as their own matched controls at a later time because they would involuntarily lapse into g Tum-mo meditation while sitting quietly for prolonged periods, as we described for subject L.T. Closely-matched, simultaneously-studied controls would have been desirable, but this was not possible in a small Himalayan hut. Despite the changing ambient temperatures and lack of between-subject control, we believe the reported results are valid for the following reasons.

It is unlikely that the reported increases in skin temperature could have occurred spontaneously. Our study was performed in relatively cold to cool ambient temperatures¹. There are two major modes of preserving body temperature in humans at cool ambient temperatures: heat conservation and increased heat production². In the first adaptation, heat flow from the body centre to the skin is reduced through vasoconstriction of the arterioles in the skin. Thus, less heat is lost to the environment, but at the expense of decreased skin temperatures. When a comfortably warm individual enters an ambient temperature of 25°C and does not change clothing or activity, his mean body skin temperature can decrease from ~33°C to 31°C, and heat loss to the environment is reduced by ~25%. Since the fingers and toes are thin cylinders with little mass in which to store heat, their temperatures usually decrease more, and much more rapidly, than those of the thorax and abdomen. Ordinarily, finger temperatures in environments such as those of our experiment are ~4°C higher than ambient³. During the meditation period, however, the subjects' finger temperatures were 10–14°C higher than ambient temperatures. The finger temperatures during the control period were already higher than expected. If the increases were spontaneous shifts, one would expect a return to normal, not a

further deviation. In the second adaptation, the metabolic rate increases, and more heat is produced. Although it is possible that these men had learned to increase their metabolism to produce more body heat, this is also unlikely. Meditation practices generally have been associated with decreased metabolic rates⁴. Furthermore, the subjects' heart rates were within normal limits and relatively stable throughout. An increased heart rate would be expected in conditions of increased metabolism.

Pasachoff extrapolated several of the increasing, control-period temperatures of subject G.J.; he disregarded those of the other subjects which were either stable or decreasing. Both Strassburg and Pasachoff mention the changing ambient temperatures as a possible confounding influence. There was, however, no consistent relationship between the temperature changes in the air and those of the skin of the subjects. For example, in G.J., the finger and toe temperature increases preceded the increases in air temperature; his finger temperature returned to baseline when ambient temperature was highest. In L.T., finger temperature increased during the meditative period at a time when ambient temperature was decreasing. As for the possibility that the location and method of attachment of the thermistors may have been partly responsible for the temperature changes, the thermistors were placed at representative skin sites of the thorax, abdomen and extremities and remained closely taped to the skin during each experiment. Both increases and decreases in skin temperature were recorded. Although there may have been some damping of the magnitude of the observed alterations in temperature, the method of application of the thermistors did not inhibit the decreases during the recovery period.

Finally, Strassburg mentions that placement of toes or fingers in an axillary cavity would bring about a rapid rise in temperature. This is true, but the subjects were under constant observation by four people throughout the experiments and assumed no such position. It is intriguing, however, to contemplate the resulting posture with one's toes in one's armpits or, perhaps, in someone else's armpits?

H. BENSON

Division of Behavioural Medicine,
Department of Medicine,
Beth Israel Hospital,
Harvard Medical School,
Boston, Massachusetts 02115, USA

1. Goldman, R. F. in *Energy Conservation Strategies in Buildings* (ed. Stolwijk, J. A. J.) (Pierce, Connecticut, 1978).
2. Berton, A. C. & Edholm, O. G. *Man in a Cold Environment* (Arnold, London, 1955).
3. VanDilla, M., Day, R. & Siple, P. A. in *Physiology of Heat Regulation and the Science of Clothing* (L. H. Newburgh, ed.) (Hafner, New York, 1968).
4. Wallace, R. K., Benson, H. & Wilson, A. F. *Am. J. Physiol.* **221**, 795–799 (1971).

BOOK REVIEWS

A kind of anthropology

R.D. Attenborough

FOR ALL his eminence, Professor Sir Edmund Leach apparently does not anticipate much critical acclaim for his contribution to Fontana's new *Master-guide* series. As he writes in the introduction, the book is not addressed to his anthropological colleagues,

many of whom are likely to be contemptuous of the style of anthropology which it advocates, and which they may well denounce as old-fashioned, egocentric, unscientific, escapist, lacking in coherence, political commitment and so on.

So readers are forewarned of an unfavourable reaction among social anthropologists, which does indeed seem to be emerging, though not entirely on the grounds predicted (see the review by Maurice Bloch, *Times Literary Supplement*, April 16). As a biological anthropologist offering a view from just outside the subject, I differ in perspective but have arrived independently at partly similar conclusions.

Leach's intention in this book is quite expressly to write about "my kind of anthropology", and accordingly the unabashed pursuit of his own intellectual concerns is evident throughout. This strongly personal flavour has its drawbacks, in that the result is far from being a balanced portrait of social anthropology as a whole, and certainly it would not make a suitable textbook for undergraduates (nor is it intended to be one). It also appears to insulate the book from criticism, since some comments are simply irrelevant to the author's aims; the deficiencies cited above are, he says, mostly intentional. But the personal flavour has its rewards too: the book is more stimulating, even entertaining and certainly more readable than would have been likely had the author adopted a blandly balanced approach. Leach leads us through his vision of social anthropology, offering at best some subtle and ingenious ideas about the understanding of human society and culture that will challenge the reader (particularly the non-specialist) to real thought; and he does this in his usual admirable manner — clear and trenchant in expression, and direct and conversational in tone.

Roughly the first half of the book deals

Social Anthropology. By Edmund Leach. Pp.254. Hbk ISBN 0-19-520371-0; pbk ISBN 0-00-635533-1. (Oxford University Press/Fontana: 1982.) Hbk £9.50, \$17.95; pbk £2.50.

with general issues, mainly different views of humanity. Leach starts by distinguishing social anthropologists from cultural anthropologists, empiricists from rational idealists, and so on. As he gives his account of these and other cross-cutting distinctions, he naturally argues for his own position — British social anthropology as a

purely taxonomic one. This first half concludes with what is perhaps a key declaration:

I myself view the varieties of human society as alternative systems of moral order rather than as a sequence of specialized adaptations to different economic circumstances.

In the second half, the focus shifts to a closer look at what social anthropologists actually study. In Leach's view, the three main spheres of concern to social anthropology are economy, kinship and cosmology (or, roughly, religion). These are not treated equally. Kinship, traditionally a prime concern in social anthropology and one which extends well beyond what readers in modern industrial society might associate with the term, is discussed at some length: the conceptual complexities of the subject, and its connections with indebtedness, power, alliance and so forth are illustrated. The interlocking relationships between economy, kinship and cosmology are also explored, but economy in its own right is not dealt with at all, and cosmology only in a few pages, despite Leach's particular interest in the subject. The book ends defiantly. Leach grants that social anthropology is "the study of the statics of social systems", despite the dynamic nature of all actual social systems; but recalling that in his earlier career as an engineer he found all the basic theory in the statics, he rests his case for the genuine value of social anthropology as he has presented it.

What, then, should one make of all this? I find a good deal to respect in Leach's cogent exposition, and up to a point I also find it persuasive. But I have reservations too. For one thing, I cannot deny a certain disappointment at the proportions of the book: the first half seems too long for its material and the second too short, especially on subjects such as cosmology. For another, the vehemence of Leach's rhetoric certainly makes for lively reading, but I think he overdoes it. The attacks on opponents are apt to become overbearing, and are sometimes plain wrong. An example of the latter comes when he accuses ethologists of assuming, among other things, that



kind of micro-sociology, blending Malinowskian empiricism and Levi-Straussian rationalism in the study of the day-to-day lives of small communities, with the aim of gaining insight, in a quasi-literary sense, into human behaviour. At the same time he denounces opposing positions, with polemical shots at conjectural historical reconstruction, cross-cultural comparison of isolated "traits", natural science thinking about human society, and other anthropological practices of which he disapproves. He then discusses criteria and concepts of humanity and animality, as a philosophical and anthropological problem rather than a

"emotions and attitudes are just as much observable characteristics as colours or structures" — a remarkable misrepresentation that presumably reflects a lack of consultation with ethologists such as his Cambridge colleague and fellow "master", Robert Hinde (whose *Ethology* was reviewed in *Nature* 296, 507; 1982).

Other attacks — for example on Geertz for an "unverifiable" statement about Balinese culture and society, on a socio-biological "just-so story", and on E.O. Wilson for the "startling incompetence" of his anthropology — may perhaps have more justice in them. The briskness with which such shortcomings are dismissed conveys the impression of commendably rigorous standards: but I think one may question whether everything Leach himself says has been scrutinized equally severely, or indeed whether it would have been reasonable to do so, given the nature of the subject matter. A possibly more straightforward issue of consistency arises within Leach's own arguments. On some points, including important ones such as the place of generalizing in social anthropology, he seems to argue in different directions as the book proceeds, creating a rather equivocal overall impression despite the clarity of particular passages.

It is conspicuous that Leach's most forceful points are frequently negative ones — not only criticizing opponents, but closing off avenues of enquiry and interdisciplinary contacts that he considers worthless. This no doubt reflects a sceptical habit of mind born of long experience, and in itself it seems defensible: indeed as an antidote to fashionable concerns it can be refreshing. Nonetheless, the cumulative effect seems to risk discouraging innovation. To my surprise, physical (biological) anthropology is one instance where Leach nods respectfully in the direction of interdisciplinary dialogue: yet this seems counter to the tenor of the argument, and the point is not pursued.

Finally, at a more mundane level, Leach is at times quite cavalier when it comes to matters of definition (e.g. species) and fact (e.g. confusing irrigation and drainage in New Guinea agriculture).

Like Aubrey Manning in his review of Hinde's *Ethology*, I am uncertain about the intended audience of the *Masterguide* series. Leach's imaginary reader is

someone like myself at the age of twenty-five to whom learning about anthropology came as an unexpected intrusion into a deceptively well-organized existence.

Social Anthropology is lively and stimulating enough to be recommended to such a reader, with the important caution that is also personal in flavour and erratic in approach. It is, for once, a book that matches its description in the cover blurb — "fascinating and idiosyncratic". □

Robert Attenborough is a Lecturer in Biological Anthropology at the Australian National University, Canberra.

Not much that's new in cancer research

Peter Alexander

Cancer: A Comprehensive Treatise, 2nd Edn. Vol. 1, *Etiology: Chemical and Physical Carcinogenesis*. Edited by Frederick F. Becker. Pp.714. ISBN 0-306-40701-9. (Plenum: 1982.) \$69.50, £43.79.

RESEARCH into cancer embraces an enormous range of disciplines, yet a broad understanding of the subject as a whole is necessary if the specialist investigator is to exploit the best opportunities available and is to evaluate the implications of his findings. A comprehensive treatment of cancer research, therefore, fills a very real need.

In 1975 such a treatise appeared. Edited by Professor F.F. Becker, the six-volume work was splendidly produced and most of the chapters were both readable and authoritative. Publication delays inherent in such a mammoth undertaking meant that the literature was only covered until 1973, and this most useful source is now only of value in some restricted areas. The pace of cancer research within the past decade has been rapid in most areas and some have changed out of all recognition. At the cellular and sub-cellular level the new techniques of gene cloning, production of monoclonal antibodies and the isolation and characterization of growth factors have provided a plethora of new facts and theories which supersede much of the earlier work.

A second edition of Becker's work was, therefore, keenly awaited and a revised version of Vol. 1 has now appeared. This, unfortunately, is a great disappointment because attempts to bring it up to date have been very patchy. The second edition contains 40 per cent more pages, but this increase is largely accounted for by the introduction of three new chapters and the extensive revision of three chapters dealing with the metabolism and biochemistry of chemical carcinogens. In all, there are 19 chapters of which six are direct reprints of the first edition with no alteration or additions; they contain no references beyond 1973 and cover the role in cancer induction of genetics, hormones, immunology and radiation. There is a further set of chapters which, while not identical reproductions of the first edition, have been changed only marginally and fail to reflect adequately the advances since 1973. Approximately half the book is, therefore, useless as a guide to the current state of knowledge. As a reference source to the older literature, these chapters have some worth, but why reprint them when copies of the first edition are readily accessible in so many libraries?

Much of the remainder is made up of some quite excellent contributions, however, which are of value both to those who wish to gain a general feel of the field

as a whole and to researchers directly engaged in the area reviewed. In general, they do not suffer greatly from the fact that, even in the new chapters, the literature is only covered until 1979. Notable amongst these are the chapters on cytogenetics by P.C. Nowell, epidemiology (V.F. Guinee), the metabolism (J.H. Weisburger and S.M. Williams) and interaction with nucleic acids (S. Rajalakshimi *et al.*) of carcinogens, and two articles on the two-stage mechanism of carcinogens by, respectively, I. Berenblum and E. Farber. What a pity that these articles were not published by themselves, since it is difficult to recommend the purchase of an expensive book in which virtually half of the contents are available in an earlier edition. □

Peter Alexander is Professor and Head of the Division of Tumour Immunology at the Institute of Cancer Research, Sutton, Surrey.

Fusion forever

R.S. Pease

Fusion. Vol.1, *Magnetic Confinement*, 2 parts. Edited by Edward Teller. Part A, pp.491, ISBN 0-12-685201-4; Part B, pp.529, ISBN 0-12-685241-3. (Academic: 1982.) Part A \$56, £37; Part B \$59, £39.

IN 1930, the installed electricity generating capacity of the Soviet Union was 2.9 GWatts. According to Professor Teller's introduction to these books, the entire supply of Leningrad — perhaps a third of the 2.9 GWatts — was offered to George Gamow for experiments to create the fusion-energy-producing temperature then recently suggested as the source of the limitless solar energy. The offer (writes Teller) was refused. Today, over 50 years later, we can anticipate such experiments as the Joint European Torus in Europe and the Tokamak Fusion Test Reactor in America coming into operation next year; they indeed use power supplies of the magnitude offered to Gamow by Bukharin. The publication of these present volumes, the first in a new series describing the state of fusion research and the hopes and uncertainties of the researchers, is thus very timely.

The first eleven chapters of the two companion volumes comprising Vol.1 are devoted to accounts of research on the confinement of hydrogen isotopes by magnetic fields. A further two chapters describe the methods proposed to heat the plasma to the required 100M°C temperatures; and the final three chapters

describe the work on technology and engineering carried out to explore the possibility of practical extraction of power from the hoped-for large-scale fusion reactions.

Each chapter is largely self-contained, and is written by a leading authority at a level which, though demanding concentration, will be rewarding for many physicists and engineers outside this field of research. At the same time, for specialists in fusion research, the articles are comprehensive, accurate, up-to-date (to about 1980) and have excellent, though not exhaustive bibliographies. Each contribution is concerned with theory and experiment in one of the main configurations of magnetic field which might form the basis of a reactor: the Tokamak, the Stellarator, the Mirror Machine, the Reverse Field Pinch and so forth. This form of presentation has the advantage that it is relatively easy to comprehend the state of play in each of these approaches, even if the reader will sometimes have to take into account what is not said as well as what is. The disadvantages are that there is some duplication of material; that the overall presentation of confinement physics is somewhat haphazard; and that important investigations of magnetic-confinement using systems, such as the Levitron, not regarded as the basis of a reactor, are omitted.

The balance of the material reflects that of the current research programme in the United States, with major articles on the theory and experiment of Tokamaks and Mirror Machines, respectively; whilst the Stellarators, though a pioneering achievement of the early US programme, are here dealt with relatively briefly. Some of the contributions bring together important material hitherto unpublished or widely scattered throughout the literature, notably the short chapter on the so-called ELMO (after St Elmo's fire) Bumpy Torus, and the chapter on Advanced Fusion Reactors. Others — those on the Tokamak system and on Magnetic Fusion Reactors — achieve a mastery over a vast literature for which both specialists and non-specialists alike will be grateful. Professor Teller's own chapter and his assembly of authors and material has ensured that there is now available for graduate engineers and physicists in general — as well as for the specialists — an up-to-date account of the main features of the physics and engineering of magnetic confinement fusion.

On the wider issue of whether or not this research will finally come to fruition, the book offers a combination of long-term confidence and short-term caution appropriate when the main ideas are subject to the imminent test of major experimentation. □

R.S. Pease is the UKAEA's Programme Director for Fusion Research, and is currently Chairman of the International Fusion Research Council.

Between the laboratory and the patient

John P. Bunker & Jinnet Fowles

Biomedical Innovation. Edited by Edward B. Roberts *et al.* Pp.395. ISBN 0-262-18103-7. (MIT Press: 1982.) \$30, £17.

MOST of us would agree that an ideal model for biomedical innovation, the process by which new diagnostic and therapeutic technologies are conceived, developed and disseminated would look something like this: a strong, adequately funded, basic research capability; prompt evaluation of the new procedure by well-designed clinical trials; a favourable balance of benefits and risks; a well-informed public, able to appreciate the consequences of treatment; timely application of the medical technology to the relief of human suffering; translation of the results into appropriate standards for general clinical practice; and continued monitoring or surveillance for long-term effects. The authors of *Biomedical Innovation* make it clear how far we are from achieving such perfection.

We are encouraged at the outset by Julius Comroe, who recounts the landmark studies in which he and the late Robert Dripps documented the benefits of strong basic science. Comroe and Dripps conducted a six-year study of the top ten clinical advances in cardiovascular-pulmonary medicine between 1945 and 1975 to determine whether they were mainly the result of targeted or of untargeted research. Their study was, in fact, a response to a Department of Defense study of weapon development which had concluded that scientists were most effective when their effort was mission-oriented. The Comroe-Dripps study, based on an examination of 4,000 published articles, came to the opposite conclusion: that the great advances in the treatment of cardiovascular disease were based on a strong, undirected, basic research effort. Since this process is already established, the message is a simple one: don't meddle.

Prompt evaluation, or the process of "knowledge validation", is described in fascinating detail by Robert Levy and Edward Sondik of the National Heart, Lung, and Blood Institute (NHLBI) — the multiple steps by which the Institute decides

which of a possible portfolio of clinical trials should be undertaken given that these trials draw funds away from the support of knowledge acquisition.

One could almost believe that the 15 per cent of its extramural budget devoted by NHLBI to clinical trials is a sufficient investment in evaluation. Unfortunately, NHLBI appears to be the exception, for ample evidence is presented elsewhere in *Biomedical Innovation* that current efforts at clinical evaluation are inadequate.

Worse, we now know that the current federal administration, as a matter of policy, intends to reduce further its investment in clinical trials in favour of support for basic research.

However unsatisfactory the record of clinical evaluation may be, the goal of a well-informed public ("informed consent") appears to be even more distant. While providing no new insights, Laurence Tancredi calls attention to four of the major unsolved conceptual problems of informed consent: the ability of the patient to comprehend the information provided by physicians; the nature of the judgements made under this uncertainty; the placebo effect; and the accuracy of available data. He concludes that these conceptual issues "create serious questions about the validity and meaningfulness of informed consent as a protective device". In his pessimistic appraisal, Tancredi slights the evidence of randomized trials demonstrating improvements in patients' understanding of the implications of treatment and even in the medical outcomes. Nonetheless, although there are indeed many obstacles to an effective informed consent process, there have been undoubted gains in the effectiveness with which information can be made available to the patient. Emphasis on the timing of information, as well as appropriate levels of vocabulary and organization, have increased patient comprehension and thoughtful decision-making.

Timely application of new medical technology is barely mentioned as an issue, which seems odd in view of the recent discussions of the so-called "drug lag" in professional journals and the lay press. Acknowledging the past existence of "considerable lag time", the editors, in an introductory chapter, state that "whether lag time exists today is essentially unanswered . . . but . . . is a concern we must continually explore". Unfortunately the reader searching for enlightenment on this thorny and controversial issue will be disappointed.

We now come to the unresolved issue which Donald Young calls "deficiencies in the communication of research results". Dr Young examines the dissemination of findings from two large clinical trials, the Veterans' Administration Cooperative Study of Surgery for Coronary Artery Disease and the NHLBI Cooperative Study of Unstable Angina Pectoris. On the basis of these two studies, he concludes that the published medical research literature is largely irrelevant to the information needs of practitioners and that there is inadequate guidance for practising physicians attempting to resolve the frequent conflicts encountered in clinical trial results. The issue he raises, the lack of evaluation of medical technology that reflects users' needs, is a theme that recurs

throughout the book — which is not surprising; in comparison with basic research, and despite their undoubted worth, evaluation studies have been woefully underfunded.

The final phase of our model, continued monitoring or surveillance, may be the most neglected. Herbert Klarman, in his chapter "Measuring Economic Effects of Biomedical Innovation", declares that monitoring the effects of technology is inherently superior to forecasting as a basis for planning. He even suggests that monitoring could serve as a basis for reimbursement of physicians and hospitals, and that this might be a better way to regulate biomedical innovation than certificate-of-need, which depends on forecasting. Great Britain has, of course, already opted for earlier introduction of drugs, relying more heavily on post-marketing surveillance. In the United States, unfortunately, and despite extensive professional debate and encouragement, no comprehensive programme of post-marketing surveillance is

yet in sight.

The method used by Young and by Levy and Sondik, as well as by several of the other authors in *Biomedical Innovation*, is "research on research" — the study of the research process as revealed in the medical literature. Much of their material has been published in other forms previously, and sadly the opportunity to use this collection to develop a cohesive and comprehensive thesis has been largely missed. But it is too much to expect a multi-authored book emerging from a three-day conference to read as if it were the product of a single, all-knowing intelligence. Rather, it is a valuable collection of essays that contains much useful information and many insights. It will be a convenient source-book for those who wish to understand the current process of biomedical innovation or who choose to write about it in the future. □

John P. Bunker and Jinnet Fowles are in the Health Services Research Program at Stanford University.

More for less

Peter W.J. Rigby

DNA Tumor Viruses, 2nd Edn revised. Edited by John Tooze. Pp.1,073. Hbk ISBN 0-87969-142-5; pbk ISBN 0-87969-141-7. (Cold Spring Harbor Laboratory: 1981.) US hbk \$75, pbk \$32.50; elsewhere hbk \$90, pbk \$39.

THIS revision of the bible of DNA tumour virology is to be warmly welcomed. That reprinting should be necessary within 18 months of publication of the original edition emphasizes the value of the book, and the editor is to be congratulated on having seized this opportunity to improve the text considerably.

The basic material is unaltered but the book has been updated by adding supplements to nine of the thirteen chapters. These supplements vary considerably in length and comprehensiveness, but they do describe the recent major advances in the field. The additional material deals effectively with such topics as polyoma middle T-antigen, the p53 protein associated with large T-antigen in SV40-transformed cells and the molecular biology of BKV, thus obviating criticisms expressed in my review of the original book (*Nature* 289, 332; 1981). The appendices of DNA sequence data have been considerably enlarged. The SV40 sequence has been modified to take account of the extra nucleotides, and another welcome addition to this section is a tabulation of restriction fragments. The sections on human adenoviruses contain a large amount of fresh sequence data and there are new appendices dealing with simian and murine adenoviruses and adenovirus-SV40 hybrids.

The book now contains some 1,100 pages which, at a price of less than \$40 for the paperback edition, represents extraordinary value. The appearance of a paperback version is the most pleasing aspect of the revision as the book is now within the economic reach of the many advanced undergraduate and graduate students who ought to own a copy. This excellent volume will be, for many years to come, compulsory reading for everyone with an interest in any aspect of DNA tumour virology. □

Peter W.J. Rigby is a member of the Cancer Research Campaign Eukaryotic Molecular Genetics Research Group and a Senior Lecturer in Biochemistry at Imperial College, University of London.

● The National Physical Laboratory has recently issued a second edition of their information poster, *Units of Measurement*. The chart summarizes the SI system and illustrates how the units are realized at the NPL. It may be obtained by writing to Information Services, NPL, Teddington, Middlesex TW11 0LW, enclosing a stamped, addressed C4 envelope.

Mulliken on molecular orbitals

W. G. Richards

Polyatomic Molecules: Results of ab Initio Calculations. By Robert S. Mulliken and Walter C. Ermler. Pp.431. ISBN 0-12-509860-X. (Academic: 1982.) \$49, £32.40.

CHARLES Coulson once suggested that the acronym MO should stand for Mulliken Orbitals rather than for Molecular Orbitals. Robert Mulliken was one of the originators of the method of obtaining solutions of the Schrödinger equation for molecules in a manner paralleling that used for atoms, each electron having its own individual function or orbital. How heartening, then, that the great man should still be active and at the forefront of the discipline. Not for Mulliken the contempt for computers; he seizes on the power which can produce accurate functions and a precision in computed properties which may on occasion rival that available from experiment.

Together with Walter Ermler, in this book he presents a comprehensive survey of *ab initio* calculations on polyatomic molecules. The epithet *ab initio* may sometimes give a spurious impression of quality; merely performing all the integrations implied in a Roothaan-Hartree-Fock calculation does not cover over defects in the original theory. In an introduction the authors do go beyond the minimum required by their title, discussing matters such as configuration interaction, electron pair and many-body approaches to improvement of wavefunction, as well as pseudopotentials.

Subsequent chapters each treat general

classes of molecules, for example AH_2 ; AB_2 ; ABC ; molecules with specified symmetry; reaction surfaces and organic compounds. In every case general considerations are followed by specific results obtained for individual molecules. Copious references mean that this is a very comprehensive survey, invaluable to theoretical chemists, to spectroscopists and to the growing band of those who describe themselves as computational chemists.

The appetite for this type of work shows no sign of abating and indeed it is being stimulated by developments in computer hardware. Cheaper and more powerful minicomputers mean that experimental chemists can run *ab initio* calculations for themselves using freely available programs obtainable from the admirable Quantum Chemistry Program Exchange run from the University of Indiana. At the same time, the new breed of supercomputer with vector and array processors offers the chance of doing extremely accurate calculations on molecules which were until recently too complex to be contemplated.

Although there is no hint of a plateau in the output of quantum mechanical calculations the book is timely in that it surveys the field, describing what has been done before the new machines have a major impact. □

W. Graham Richards is a Lecturer in the Physical Chemistry Laboratory at Oxford University. The latest volume of his Bibliography of *Ab Initio* Molecular Wavefunctions covering 1978-1980 was published by Oxford University Press last year.